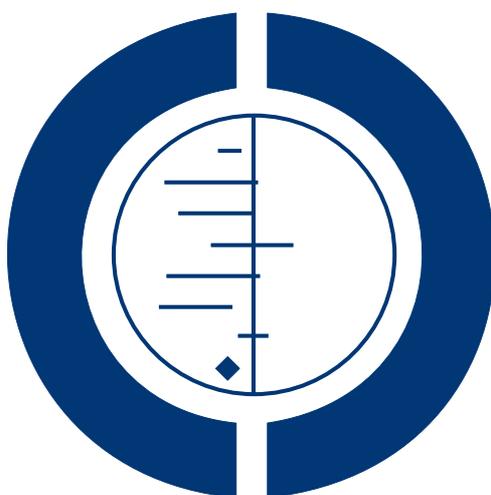


# Technical editing of research reports in biomedical journals (Review)

Wager E, Middleton P



**THE COCHRANE  
COLLABORATION®**

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2008, Issue 4

<http://www.thecochranelibrary.com>



## TABLE OF CONTENTS

HEADER . . . . .	1
ABSTRACT . . . . .	1
PLAIN LANGUAGE SUMMARY . . . . .	2
BACKGROUND . . . . .	2
OBJECTIVES . . . . .	3
METHODS . . . . .	3
RESULTS . . . . .	4
DISCUSSION . . . . .	8
AUTHORS' CONCLUSIONS . . . . .	11
ACKNOWLEDGEMENTS . . . . .	12
REFERENCES . . . . .	12
CHARACTERISTICS OF STUDIES . . . . .	17
DATA AND ANALYSES . . . . .	51
APPENDICES . . . . .	64
WHAT'S NEW . . . . .	64
HISTORY . . . . .	65
CONTRIBUTIONS OF AUTHORS . . . . .	65
DECLARATIONS OF INTEREST . . . . .	65
SOURCES OF SUPPORT . . . . .	66
INDEX TERMS . . . . .	66

[Methodology Review]

# Technical editing of research reports in biomedical journals

Elizabeth Wager<sup>1</sup>, Philippa Middleton<sup>2</sup>

<sup>1</sup>Sideview, Princes Risborough, UK. <sup>2</sup>ARCH: Australian Research Centre for Health of Women and Babies, Discipline of Obstetrics and Gynaecology, The University of Adelaide, Adelaide, Australia

Contact address: Elizabeth Wager, Sideview, 19 Station Road, Princes Risborough, Buckinghamshire, HP27 9DE, UK. [liz@sideview.demon.co.uk](mailto:liz@sideview.demon.co.uk).

**Editorial group:** Cochrane Methodology Review Group.

**Publication status and date:** New search for studies and content updated (no change to conclusions), published in Issue 4, 2008.

**Review content assessed as up-to-date:** 25 July 2007.

**Citation:** Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database of Systematic Reviews* 2008, Issue 4. Art. No.: MR000002. DOI: 10.1002/14651858.MR000002.pub3.

Copyright © 2008 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

### Background

Most journals try to improve their articles by technical editing processes such as proof-reading, editing to conform to 'house styles', grammatical conventions and checking accuracy of cited references. Despite the considerable resources devoted to technical editing, we do not know whether it improves the accessibility of biomedical research findings or the utility of articles. This is an update of a Cochrane methodology review first published in 2003.

### Objectives

To assess the effects of technical editing on research reports in peer-reviewed biomedical journals, and to assess the level of accuracy of references to these reports.

### Search strategy

We searched The Cochrane Library Issue 2, 2007; MEDLINE (last searched July 2006); EMBASE (last searched June 2007) and checked relevant articles for further references. We also searched the Internet and contacted researchers and experts in the field.

### Selection criteria

Prospective or retrospective comparative studies of technical editing processes applied to original research articles in biomedical journals, as well as studies of reference accuracy.

### Data collection and analysis

Two review authors independently assessed each study against the selection criteria and assessed the methodological quality of each study. One review author extracted the data, and the second review author repeated this.

### Main results

We located 32 studies addressing technical editing and 66 surveys of reference accuracy. Only three of the studies were randomised controlled trials.

A 'package' of largely unspecified editorial processes applied between acceptance and publication was associated with improved readability in two studies and improved reporting quality in another two studies, while another study showed mixed results after stricter editorial

policies were introduced. More intensive editorial processes were associated with fewer errors in abstracts and references. Providing instructions to authors was associated with improved reporting of ethics requirements in one study and fewer errors in references in two studies, but no difference was seen in the quality of abstracts in one randomised controlled trial. Structuring generally improved the quality of abstracts, but increased their length. The reference accuracy studies showed a median citation error rate of 38% and a median quotation error rate of 20%.

### **Authors' conclusions**

Surprisingly few studies have evaluated the effects of technical editing rigorously. However there is some evidence that the 'package' of technical editing used by biomedical journals does improve papers. A substantial number of references in biomedical articles are cited or quoted inaccurately.

## **PLAIN LANGUAGE SUMMARY**

### **Technical editing of articles before they are published in medical journals.**

Most journals try to improve articles before publication by editing them to make them fit a 'house-style', and by other processes such as proof-reading. We refer to all these processes as technical editing. We identified 32 studies of the effects of technical editing from a systematic review. There is some evidence that the overall 'package' of technical editing raises the quality of articles (suggested by 'before-and-after' studies) and that structuring abstracts makes them more useful, although longer. However, there has been little rigorous research to show which processes can improve accuracy or readability the most, or if any have harmful effects or disadvantages.

Over one third of references cited in articles in medical journals have some inaccuracies and one-fifth of quotations to references in these articles are not accurate

## **BACKGROUND**

The contents of peer-reviewed journals should be accurate and complete and should present research findings in a responsible and comprehensible way. Since healthcare workers often have little time to read papers it is also beneficial if the material can be read both quickly and correctly (Sackett 2000). The presentation of research in peer-reviewed journals should therefore not confuse or mislead even if the reader has time only to scan the text. Journals try to maximise the accessibility, completeness and accuracy of information by specifying the format and style of papers that are acceptable (e.g. in the Instructions to Contributors) and by performing checks on accepted material (e.g. proof-reading). Most journals also make an active attempt to 'improve' the presentation of papers and ensure they conform to 'house style' and grammatical conventions through the process of copy-editing. Authors are also involved in formatting papers and proof-reading (i.e. checking the version intended for publication against the original to identify typographic errors and checking that changes introduced during copy-editing are acceptable). We shall refer to these processes collectively as technical editing.

Despite the time and resources devoted to layout and technical editing most journals do not present any evidence about the effects of their design (e.g. type face, column width) or their house style (e.g. use of abbreviations, presentation of numbers) on legibility, readability or comprehension, nor do they measure the effects of technical editing on the ability of readers to draw valid conclusions from papers (Overbeke 1999). This review examines the evidence of the effects of technical editing processes undertaken by biomedical journals on original research papers and those performed by the papers' authors after a paper has been accepted for publication. The review also examines the evidence of the effects of journals' house styles and recommendations for data presentation on published papers.

Two main aspects of technical editing will be considered: the effects of different journal styles and the methods for ensuring compliance with these styles. These need to be considered separately in order to distinguish the effects of imposing an inappropriate style (e.g. one that reduces readability or comprehension) from ineffective methods of achieving compliance with a 'good' style (i.e. one that

improves accessibility).

## OBJECTIVES

To assess the effects of technical editing performed on papers accepted for publication by peer-reviewed biomedical journals on the papers' accuracy, consistency, completeness, legibility, readability, comprehensibility or other measures thought to reflect or influence the usefulness of the paper to the reader or its effects on the reader's knowledge, attitude or behaviour. The appropriateness of outcome measures are reviewed and the costs of technical editing, in terms of journals' and authors' time and resources, are assessed where possible. We also assessed the level of accuracy of references to these reports.

The review focuses on processes designed to correct genuine mistakes rather than those aimed at detecting scientific fraud, and will concentrate on the presentation of research findings rather than their generation. (Methods assessing research validity, methodological and ethical soundness, etc. are addressed in another review on editorial peer review (Jefferson 2007).

This is an update of a Cochrane methodology review first published in 2003. Despite more than doubling the number of included studies, the conclusions remain largely unchanged.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

Prospective or retrospective comparative studies with two or more comparison groups were included. All studies had to report original data. These groups could be generated by random or other methods and could include historical comparisons.

Non-comparative studies were included to provide information on 'background' levels (e.g. of citation accuracy in published papers) to provide an estimate of current quality and to stimulate research into interventions designed to raise this.

#### Types of data

Evidence was reviewed from studies relating to original research articles published in biomedical journals. When studies related to readability or comprehension they had to include participants drawn from the usual readership of the journal (i.e. healthcare professionals). Papers in any language were considered but evidence about writing style (e.g. use of passive voice, sentence length) needed to relate primarily to studies of English-language publications.

#### Types of methods

Studies comparing two or more interventions or an intervention against doing nothing from within one of the following categories were included:

- differences in / absence of instructions to contributors / authors
- differences in journal house style and page layout
- different methods of data presentation
- imposition of quantifiably-different writing styles (e.g. passive voice, sentence length)
- copy-editing
- proof-reading

#### Types of outcome measures

Accuracy, completeness, consistency, legibility, readability (e.g. Gunning Fog Index and Flesch Reading Ease Score) and comprehensibility of the published report, however measured. Other measures that influence or reflect the usefulness of the published report to the reader or the cost of the technical editing process. An interpretation of the levels of difficulty in Flesch Reading Ease and Gunning Fog Index scores is given in [Table 1](#).

**Table 1. Interpretation of readability indexes**

Flesch score	Gunning Fog index	Description	Example
90-100	5	very easy	Reader's Digest
80-89	6	fairly easy	Time
70-79	7	easy	US News

**Table 1. Interpretation of readability indexes** (Continued)

60-69	9	standard English	New York Times
50-59	12	fairly difficult	The Ambassadors, by Henry James
30-49	14-16	difficult	Corporate annual report
0-29	16	very difficult	legal contract

### Search methods for identification of studies

We searched the following databases: Cochrane Methodology Register, MEDLINE, and EMBASE. For the full search strategy, see [Appendix 1](#).

Reference lists from [Godlee 1999](#) were searched.

Reference lists of retrieved relevant articles were searched.

Simple search strategies were used, since technical editing topics appear to be inconsistently indexed in most databases. The search strategy for each database generally consisted of a single concept term (text as well as controlled language if available). This resulted in poor specificity but was felt necessary in order to obtain optimal sensitivity in locating relevant studies.

Some search terms retrieved no or very few relevant citations and so these searches were not repeated in other databases (e.g. authorship on MEDLINE).

### Data collection and analysis

Two review authors (EW and PM) independently examined each retrieved citation. Those thought to fulfil the selection criteria were retrieved in full. Two review authors (EW and PM) compared each article against the selection criteria independently, resolving disagreements by discussion.

In the first version of the review, one review author (EW) extracted data on the effects of technical editing, recording the study design and assessing the appropriateness of outcome measures. A second review author (PM) then repeated the data extraction. This process was reversed for the 2007 update. Thus data extraction was sequential rather than truly independent but was done without reference to the first extraction. A small number of minor discrepancies were resolved by discussion. Study authors were contacted for missing data or for clarification.

None of the included studies were regarded to be sufficiently similar in purpose, design, methodology or outcomes to combine statistically. However we did calculate median reference accuracy

rates across journals. It was not possible to make any assessments of statistical heterogeneity, or even of general heterogeneity. Neither was it possible to carry out any subgroup or sensitivity analyses, or to make any formal assessment of the likelihood of publication bias.

Gunning Fog and Flesch scores are often used to measure readability, but their ability to do this validly and reliably has been questioned. However they may give some idea of relative difficulty of versions of the same text ([Hartley 2000a](#)). Both systems produce scores based on sentence and word length, but do not take account of word familiarity, sentence complexity or individual differences in perceptions of text difficulty ([Connatser 1999](#)) and so can be considered, at best, surrogate markers of comprehensibility.

## RESULTS

### Description of studies

See: [Characteristics of included studies](#); [Characteristics of excluded studies](#).

We identified 32 studies addressing four aspects of technical editing, and which fulfilled our other inclusion criteria. We also identified an additional 69 studies which described the accuracy of references in peer-reviewed biomedical journals, and included 66 of these. (These are annotated with RA at the end of each study identifier e.g. [Asano 1995a RA](#)). Eight of the 69 reference accuracy studies ([Asano 1995b RA](#); [George 1994 RA](#); [Hobma 1992 RA](#); [Jackson 2003 RA](#); [Lowry 1985 RA](#); [Nishina 1995c RA](#); [Nishina 2000 RA](#); [Oermann 2002b RA](#)) were also included in the technical editing category, since they contained some data about interventions undertaken to improve accuracy. See also the [Characteristics of included studies](#) table.

Eight studies (seven reference accuracy and one study of technical editing) are awaiting classification. We understand that the technical editing study is unlikely to be fully published.

## I. TECHNICAL EDITING

### I.1 PEER REVIEWING AND EDITING REPORTS (14 studies)

Five studies examined the impact of peer reviewing and editing on submitted manuscripts by measuring readability or reporting quality (Biddle 1996; Goodman 1994; Laccourreye 1999; Pierie 1996; Roberts 1994).

Pitkin 1999; Pitkin 2000; Silagy 1998; and Winker 1999 all compared abstracts before and after quality improvement initiatives or specialist editing.

In George 1994 RA; Hobma 1992 RA; Lowry 1985 RA; and Oermann 2002b RA reference accuracy was compared before and after some form of editorial review; and Siegel 2005 compared the information content of titles from journal articles.

### I.2 PROVIDING INSTRUCTIONS TO AUTHORS (7 studies)

Karlawish 1999 compared the quality of reporting research ethics with the amount of detail provided in authors' instructions in 21 journals.

Pitkin 1998 used a randomised controlled trial to test the hypothesis that providing authors with specific instructions results in more accurate abstracts.

The Fister 2005 randomised trial compared an instructional intervention, and a brief reminder compared with standard practice, to see if these prompted authors to reduce the numbers of errors in references in manuscripts submitted to a journal; and Asano 1995b RA; Jackson 2003 RA; Nishina 1995c RA; Nishina 2000 RA surveyed citation errors before and after asking authors to check their references.

### I.3 PROVIDING INSTRUCTIONS TO READERS (1 study)

Gross 1994 compared the problem-solving ability of journal readers provided with an enhanced article, with that of readers provided with the original article.

### I.4 STRUCTURING ABSTRACTS (18 studies)

A large number of studies assessed the effects of structuring of abstracts on:

- reporting quality of the abstract (Dupuy 2003; Khosrotehrani 2002; Scherer 1998; Taddio 1994; Trakas 1997; Wong 2005);
- readability, length and/or format of the abstract (Comans 1990; Hartley 1997; Hartley 1998; Hartley 2002; Hartley 2003);
- accuracy of the abstract (Hartley 2000);
- reader preference (Hartley 1996c);

- retrievability of the abstract (Booth 1997; Harbourt 1995; Hartley 1996a; Hartley 1996b; Wilczynski 1995).

## 2. REFERENCE ACCURACY

The majority of studies found were 'baseline surveys' which measured reference accuracy at one time point. A handful of studies made comparisons between journals or between different years of a journal (see comments column in *Analysis 2.1*), but most made no attempt to link their findings to specific editorial interventions. One study (Riesenberg 2001 RA) was a review of reference inaccuracies from 30 studies published between 1979 to 2000 in the biomedical literature, including dentistry, nursing, medicine, pharmaceutical, public health, science, and veterinary medicine. Several investigators have examined two aspects of reference accuracy:

(1) Citation accuracy measures the accuracy of the reference list and checks that details such as the authors' names, date of publication, journal name, volume and page numbers are correct by comparing them with the original source or an authoritative database such as MEDLINE

(2) Quotation accuracy involves more subjective tests to see whether findings from other studies or statements by other authors are accurately reflected in the papers citing these 'quotations'.

A major quotation error was generally defined as a seriously misleading change to the original quotation and a major citation error was generally defined as one that prevented or seriously obstructed the identification or retrieval of the reference, e.g. incorrect volume number. A minor citation error was one that did not prevent readers from retrieving the citation, e.g. misspelling of an author's name.

## Risk of bias in included studies

## I. TECHNICAL EDITING

Only three of the 32 technical editing studies (Fister 2005; Gross 1994; Pitkin 1998) were randomised trials. Fister 2005 assessed the effect of a brief reminder or an instructional intervention on reference accuracy; Gross 1994 assessed advice to authors on how to present data and Pitkin 1998 assessed the impact of printed instructions to authors on abstract quality.

The remainder of the studies were either:

- prospective comparative studies (Booth 1997; Comans 1990; Dupuy 2003; Harbourt 1995; Hartley 1996a; Hartley 1996b; Hartley 1996c; Hartley 2000; Hartley 2002; Hartley 2003; Siegel 2005; Wilczynski 1995);
- retrospective comparative studies (Karlawish 1999; Pitkin 1999);
- before-and-after studies (Biddle 1996; Goodman 1994; Pierie 1996; Wong 2005 - blinded; Laccourreye 1999; Roberts

1994; Scherer 1998 - probably not blinded; Taddio 1994; Trakas 1997 - partially blinded; Hartley 1998; Khosrotehrani 2002; Pitkin 2000; Silagy 1998; Winker 1999 - not blinded).

- Hartley 1997 was a set of eight studies, some before-and-after and some comparative studies.

Eight reference accuracy studies were also included in the technical editing category. These were all retrospective comparative studies (Asano 1995b RA; George 1994 RA; Hobma 1992 RA; Jackson 2003 RA; Lowry 1985 RA; Nishina 1995c RA; Nishina 2000 RA; Oermann 2002b RA).

Most studies were small and probably under-powered, with the potential for confounding effects.

## 2. REFERENCE ACCURACY

The majority of these studies were 'snapshots' of error rates in one or a small number of journals, meaning extrapolation of results over time, or extrapolation to larger groupings of journals may not be reliable.

### Effect of methods

## 1. TECHNICAL EDITING

### 1.1 PEER REVIEWING AND EDITING REPORTS (14 studies)

#### Summary:

A combined 'package' of peer-review and editorial processes improved readability in two studies (Biddle 1996; Roberts 1994) and reporting quality was improved in another study (Goodman 1994). Silagy 1998 found that abstracts of Cochrane reviews which had been professionally edited by the journal Evidence-Based Medicine were clearer and more consistent than the original abstracts. Pierie 1996 measured the effects of peer-review and editorial processes separately and while both interventions improved manuscript reporting quality, these improvements were in different areas. One study (Laccourreye 1999) found that 12 aspects significantly improved, four aspects (including the number of errors per page) worsened, and there was no apparent change in 20 other measures of quality following the introduction of stricter editorial policies. Editorial processes appeared to lessen the numbers of errors in abstracts (Pitkin 2000; Winker 1999) and references (George 1994 RA; Hobma 1992 RA; Lowry 1985 RA; Oermann 2002b RA). Pitkin 1999 found significant differences between journals in regard to the proportion of deficient abstracts but did not speculate on the cause of this. Siegel 2005 found the BMJ was the only one of four leading medical journals to show an increase over time in the number of articles with titles that included information about methods used in the study.

*Detail (also see Analysis 1.1):*

In Biddle 1996, while peer-review and editorial processes significantly improved readability scores, readability remained in the 'difficult' category. Both computerised and manual scoring produced the same finding. Articles were significantly shorter after peer review and editing.

In Goodman 1994, 33 out of 34 items in a quality assessment instrument showed improvement after peer-review and editing processes, though only four of these showed a statistically significant change ( $p = 0.05$  or less).

Laccourreye 1999 compared the quality of scientific reports over time and found a decrease in the quality of titles but an increase in the quality of materials and methods, and results sections. No change in quality was detected for summaries, introductions and discussions. The median number of errors showed a statistically significant increase over time.

Pierie 1996 looked at peer review (comparison of submitted and accepted versions) and editing (comparison of accepted and published versions). Fourteen out of 23 questions (61%) about manuscript quality showed significant improvement ( $p=0.03$  or less) after peer review. There was also a significant difference in the overall score (ratings of three or more on a five-point scale) with the score improving from 59% to 81% after peer review, 22% difference (95% CI 15.0 to 27.1),  $p=0.00001$ . Eleven out of 16 questions about manuscript quality (69%) showed significant improvement ( $p=0.017$  or less) after editing, especially in style and readability. Questions were structured on a five-point scale as before.

In Roberts 1994, peer-review and editorial processes significantly improved readability, although the scores remained in the highest categories of difficulty.

Pitkin 1999 found that deficient abstracts were common (39%; 104/264) in six general medical journals, although proportions varied widely (and statistically significantly) from 18% to 68% between the journals. A deficient abstract was defined as one that was inconsistent with the text of the article and/ or contained material not found in the text.

After a quality improvement initiative consisting of 11 criteria "developed using evidence wherever possible, built on work of previous authors and supplemented by common sense" (Winker 1999) at JAMA, Pitkin 2000 found that the number of overall deficiencies in abstracts dropped from 26/50 (52%, 95% CI 38% to 66%) to 10/50 (20%, 95% CI 9% to 31%), chi-square = 11.11,  $p<0.005$ .

In a similar study to Pitkin 2000, Winker 1999 found that the number of deficient abstracts in JAMA dropped after the introduction of an editorial quality improvement initiative.

George 1994 RA (references) found that the relationship between rate of citation errors and journals that monitor citations, was of "borderline significance".

Hobma 1992 RA (references) found that references in submitted articles were less accurate than published references (i.e. after ref-

ences had undergone editorial scrutiny).

Lowry 1985 RA (references) found more inaccurate quotations and citations in the correspondence received than the correspondence published in BMJ in the same time period.

## 1.2 PROVIDING INSTRUCTIONS TO AUTHORS (7 studies)

### Summary:

One non-randomized study (Karlawish 1999) suggested that improved reporting quality in the area of research ethics was associated with journals that provided more detailed instructions to authors. Four before-and-after studies of the effect of requiring authors to supply photocopies of the first page of their references (Asano 1995b RA; Jackson 2003 RA), and of instructing authors to check original sources (Nishina 1995c RA; Nishina 2000 RA), showed an increase in citation accuracy. In a randomised trial, Fister 2005 also found that instruction provided by the journal editors to authors resulted in small improvements in the accuracy of references. In another randomised trial, Pitkin 1998 was unable to detect any difference in the quality of abstracts prepared by authors who had received instructions about preparation of abstracts, compared to authors who did not receive these instructions. Twenty-eight percent (25/89) of abstracts in the instructed group contained defects (95% CI 19% to 37%) compared with 25% (30/114) in the uninstructed group (95% CI 18% to 34%),  $p = 0.78$ . See Analysis 1.1 for more detail.

## 1.3 PROVIDING INSTRUCTIONS TO READERS (1 study)

In Gross 1994, the inclusion of a clinical example in a research report made no discernible difference to the ability of therapists to select the correct model to predict knee function for different patients. However those therapists randomized to receive the worked example showed a greater ability to solve a clinical problem mathematically (i.e. calculating predicted muscle performance),  $\chi^2 = 9.35$  ( $df=1$ )  $p < 0.01$ .

## 1.4 STRUCTURING ABSTRACTS (18 studies)

### Summary:

While a sample of structured abstracts compared with an unpaired sample of unstructured abstracts showed no difference in readability scores (Hartley 1997), unstructured abstracts rewritten as structured abstracts showed improved readability scores (Hartley 1997; Hartley 1998; Hartley 2003) as well as several other measures of comprehensibility in Hartley 2003. Structured abstracts were longer than unstructured abstracts (Comans 1990; Dupuy 2003; Harbourn 1995; Hartley 1997; Hartley 1998; Hartley 2003) but this extra length can usually be accommodated without increasing the overall number of pages for the article (Hartley 2002). Readers were able to unscramble structured abstracts more easily than unstructured abstracts (Hartley 1997) and they preferred

structured abstracts (Hartley 1997). In a study by Hartley 1996c, readers' preferred layout of structured abstracts included bold capitals for subheadings, a line space between headings and for the abstract to run across the page in a single column as opposed to being in a two column format. The reporting quality of structured abstracts was better than for unstructured abstracts in four studies (Dupuy 2003; Taddio 1994; Trakas 1997 and Wong 2005) with no difference seen in a fifth study (Khosrotehrani 2002). In a study of randomised trials using CONSORT criteria (Scherer 1998), structured abstracts were of higher quality in one journal, but no difference between structured and unstructured abstracts was seen in a second journal and no difference was detected in the reporting quality of the text for either journal. Hartley 2000 found few differences between the accuracy of unstructured and structured abstracts in the psychological literature. Booth 1997 found that electronic searching precision (specificity) may be a little better with structured abstracts, but recall (sensitivity) was probably worse for structured abstracts compared with unstructured abstracts. In Wilczynski 1995 many search terms were comparable for structured and unstructured abstracts, but some performed better in MEDLINE with structured abstracts, particularly for aetiology and prognosis articles. Harbourn et al's study of MEDLINE records (Harbourn 1995) found that structured abstracts had more access points (in the form of Medical Subject Headings) than the overall sample of MEDLINE records. Two studies (Hartley 1996a; Hartley 1996b) indicate that readers find it easier to search structured abstracts than unstructured ones.

*Detail (also see Analysis 1.1):*

### Comans 1990

When 10 unstructured abstracts from *Nederlands Tijdschrift voor Geneeskunde* were rewritten as structured abstracts, their length increased from a mean of 163.7 words (SD 41.6) to 263.6 words (SD 52.6).

### Dupuy 2003

In a comparison of abstracts of clinical studies published in 2000 in three dermatology journals, structured abstracts ( $n=34$ ) scored significantly better than unstructured abstracts ( $n=15$ ): 0.71 (SD 0.11) versus 0.56 (SD 0.18);  $P=0.002$ . Structured abstracts were longer on average than unstructured abstracts: 256 words (SD 77) versus 169 (SD 65);  $p < 0.001$ . The scoring system was adapted from Narine 1991. A strong positive correlation between length and score was observed for unstructured abstracts ( $p=0.002$ ) while no such significant correlation was seen for structured abstracts ( $p=0.08$ ).

### Harbourn 1995

Structured abstracts were on average 700 characters longer and had three more Medical Subject headings than MEDLINE records as a whole.

### Hartley 1998

Abstracts showed overall improvement in their readability when the original unstructured abstract was rewritten by the original

author as a structured abstract:  $t = 2.81$ ,  $p$  (one-tailed)  $<0.005$  for Flesch Reading Ease and  $t = 3.77$ ,  $p$  (one-tailed)  $<0.0005$  for Gunning Fog Index. The structured abstracts were longer, with unstructured abstracts having a mean of 147.4 words (SD 47.5) and structured abstracts having a mean of 210.6 words (SD 50.5),  $t = 8.54$ ,  $p$  (one-tailed)  $<0.0005$ .

[Hartley 1997](#)

Using readability formulae, Hartley and Sydes found that structured abstracts do not appear to be any easier to read than unstructured abstracts, but versions rewritten as structured abstracts did score better. In this study, readers found some scrambled structured abstracts easier to reconstruct than scrambled unstructured ones, particularly when the style and the presentation of the abstracts differed a great deal. Readers rated structured abstracts as easier to read than unstructured ones.

While [Wong 2005](#) found an improvement in quality when structured abstracts were introduced, no such increase in quality was seen over time for the structured abstracts.

## 2. REFERENCE ACCURACY

### Summary:

Over 27,000 references have been checked in accuracy studies in the biomedical literature, with 6,962 out of 23,313 (30%) references having at least one citation error. For quotation errors, 761 out of 3836 (20%) references were quoted inaccurately. The median citation error rate per journal was 38%, with a range of 4% to 67%. The median quotation error rate per journal was 20%, with a range of 0% to 50%. (Because of its different methodology, the results from Kolbitsch 1997 RA could not be included in this calculation.)

Detailed results for each study are shown in [Analysis 2.1](#).

[Riesenberg 2001 RA](#) reviewed 30 studies of reference accuracy published between 1979 and 2000, finding citation error rates of 7% to 60% (with 1% to 24% major errors) and quotation errors of 0% to 58%. The results from this review were not included in the above calculations of median error rates as this would have double-counted studies we had already included in this review.

Although the [Fister 2005](#) randomised trial examined the accuracy of references, we felt we could not use their prevalence data in the reference accuracy table, as their definitions of reference accuracy were markedly different from the other studies of reference accuracy.

## DISCUSSION

Remarkably little research into the effects of editing performed by biomedical journals has been published. The literature contains a large volume of opinion and discussion, not much evidence, and even fewer rigorous studies. The biggest limitation in interpreting the included studies is the inability to make many valid qualitative

or quantitative comparisons between them, due to diversity in design and outcomes, but more fundamentally in topics.

Those studies that have been published fall into four broad categories: those measuring the effects of the total (sometimes unspecified) package of editing that occurs between submission or acceptance and publication; those examining the effect of providing authors with instructions; those measuring the effect of structuring abstracts; and those considering the accuracy of cited references.

## EFFECTS OF THE ENTIRE EDITORIAL 'PACKAGE'

### (i) Effects of editing on overall manuscript quality

[Pierie 1996](#) (at the *Nederlands Tijdschrift Geneeskunde*) and [Goodman 1994](#) (at the *Annals of Internal Medicine*) both investigated these journals' normal editorial processes on manuscript quality by measuring the changes that occur between manuscript submission and publication, and both reported that published versions of manuscripts received higher quality scores than submitted versions. Pierie also found that different improvements were introduced between submission and acceptance, and between acceptance and publication. Presumably the submission/acceptance comparison highlights the peer-review process, and the acceptance/publication comparison highlights the in-house editing (including technical editing) processes. [Pierie 1996](#) explains that "During editing, the information in the article is checked scientifically and linguistically, corrected and clarified if necessary, numbers are checked when possible, and the references are made to conform to the so-called Vancouver system". In contrast, the [Goodman 1994](#) study measured the combined effect of peer review, editors' comments and technical editing.

In both [Pierie 1996](#) and [Goodman 1994](#), the investigators developed and used their own, non-validated scoring systems so they cannot be directly compared or pooled, although in both cases the improvements were small. The studies also recruited different types of assessors; 'expert' in [Goodman 1994](#) and volunteer readers in [Pierie 1996](#). While both studies were before-and-after designs and both masked the different versions, [Goodman 1994](#) may have been less biased because different evaluators assessed different versions. However, this may have contributed to the low reliability of the assessment instrument used in [Goodman 1994](#). Over 70% of the assessors in the [Pierie 1996](#) study correctly identified the three versions they received (i.e. masking was unsuccessful) and those who correctly identified the published version gave significantly higher scores than those who failed to recognise it.

It is difficult to comment on whether the results from these two studies might also apply to other journals, although [Goodman 1994](#) comments that "the relatively large editorial staff at *Annals* is not typical of any but the largest medical journals, and the

generalization to others with different selection, review and editing processes cannot easily be made”.

Laccourreye 1999 is also a before-and-after study, although the interventions are not clearly specified. In 1990 the journal's editorial policy became stricter; “les editeurs ... demandant aux auteurs et aux experts un respect strict des regles de redaction de l'article original” [the editors asked the authors and reviewers to follow strict rules for editing original papers], but the introduction of bias from confounding (e.g. release of the Uniform Requirements for Manuscripts submitted to Biomedical Publications in 1991), and from the difficulty of masking the papers in a single-author study, is quite likely. The improvements over time, and those attributed to new editorial policies, occurred from a low baseline (e.g. 30% of original research papers published in 1977 contained no references and only 44% followed the IMRAD (Introduction Methods Results and Discussion) structure) and so these results are not likely to be generalisable to other journals.

#### (ii) Effects of editing on readability

Both studies that examined the effects of editing on readability found improvements in readability scores, although papers remained difficult to read. Neither study author (Biddle 1996; Roberts 1994) provided details or comment about which aspects of editing were thought to affect readability. As outlined in the Methods section, Gunning Fog and Flesch scores may not be reliable measures of readability.

#### (iii) Effects of editing on quality of abstracts

Two small before-and-after surveys (Pitkin 2000; Winker 1999) found that more rigorous in-house editing resulted in abstracts with fewer deficiencies. An earlier survey (Pitkin 1999) had found that deficient abstracts were common and that there were significant differences in the proportions of defective abstracts between six unnamed journals, which may reflect different editorial practices. Silagy 1998's findings of the impact of professional editing of abstracts of Cochrane reviews may no longer apply since The Cochrane Collaboration now devotes more attention to abstract quality. However the study does demonstrate the contribution that professional technical editing can make to the quality of scientific writing.

#### (iv) Effects of editing on quality of references

In Lowry 1985 RA, nearly all references published (92%) were accurate compared with 69% of references in letters submitted to the *BMJ*. Lowry attributes this difference to the checking by subeditors who correct any obvious errors during the process of putting references into the house style “which allows many mistakes to be spotted, especially where the fault is an incomplete reference, which is inevitably corrected”. In this small study, the two

groups of letters may have differed in other ways apart from their publication status and the study author was aware of the status of each letter.

In another small open survey, Hobma 1992 RA also found that papers published in the *Nederlands Tijdschrift voor Geneeskunde* contained more accurate references (69%) than did submitted papers (30%), and that 11 of the 31 inaccurate references that were published “could not have been prevented in the [normal] editorial process because they were listed wrongly or not at all in Index Medicus”.

In George 1994 RA, the combined reference error rate of the journals that monitored citations was lower than in those that did not (35% versus 48%), the difference being “of borderline significance  $p = 0.066$ ”. However the sample may have been too small to detect a significant difference between the journals.

There was some indication that major journals and journals with higher impact factors had a lower error rate than 'minor' journals (Lok 2001 RA; Warren 1997 RA) and this may be attributable to major journals have more access to editing resources. This is consistent with the observation in Oermann 2002b RA that journal librarian checks of references helped to improve the accuracy of references. In the future, computer software may be able to link incorrectly cited references with the master version and make automatic corrections. At present, the ability to link electronically MEDLINE references, for example, should lead to lower numbers of citation errors in references, but this will not address the question of quotation errors.

## PROVIDING INSTRUCTIONS TO AUTHORS

We found only one study (Karlawish 1999) that investigated the effect of providing different levels of author instructions on the quality of reporting. Although a positive effect was reported when more detailed instructions were provided, the study was small, the survey design would not have precluded confounding and the specialised nature of the study (instructions to authors writing about topics involving research ethics in nursing home settings) may not be generalisable to other topics and settings. The more rigorous randomised trial design of Pitkin 1998 found no evidence of an effect for sending specific instructions to authors although the Fister 2005 randomised trial found small improvements in the accuracy of references. The type of instructions provided (in terms of content or presentation) and the 'passive' dissemination method may have also influenced the findings. Four before-and-after studies asked authors to verify their references, by supplying photocopies of the first page of their references (Asano 1995b RA; Jackson 2003 RA) and asking authors to check original sources (Nishina 1995c RA; Nishina 2000 RA). More accurate citation of references was reported in both studies after the interventions, but again other factors may have had a confounding effect and the

investigators (probably aware of the year of publication) may have unwittingly applied tougher criteria to the older references.

## INTERVENTIONS DESIGNED TO IMPROVE QUALITY BY STRUCTURING ABSTRACTS

In a before-and-after study, [Taddio 1994](#) found that structured abstracts from three journals rated more highly than unstructured abstracts from the same journals (prior to the introduction of the structured format). Raters were masked to the identity of the journal (but it was obviously not possible to mask the type of abstract) and inter-rater agreement was high. There is a possibility of confounding, but the findings were consistent across the three journals. Similar conclusions were drawn by [Trakas 1997](#) who considered abstracts of pharmacoeconomic studies using criteria adapted from [Taddio 1994](#). [Trakas 1997](#) found that medical journals tended to use structured abstracts while health economics journals did not, so abstract quality may have been affected by other factors relating to particular journals.

[Scherer 1998](#) found some fairly weak indications that structured abstracts fared better than unstructured abstracts on the CONSORT criteria, but their study is likely to have been underpowered ([Mago 1999](#)). In addition, the CONSORT statement was very new and therefore was unlikely to have made an impact on the quality of reporting of either structured or unstructured abstracts. A recent systematic review of eight studies has concluded that adoption of CONSORT by journals is associated with improved reporting of randomized trials ([Plint 2006](#)).

The rewriting of 10 unstructured abstracts from the *Nederlands Tijdschrift voor Geneeskunde* into structured abstracts by a single investigator ([Comans 1990](#)), made the abstracts longer. Unstructured abstracts were also rewritten in two other studies ([Hartley 1997](#); [Hartley 1998](#)). Like [Comans 1990](#), some of the abstracts in the [Hartley 1997](#) study were rewritten by the investigators, while in the [Hartley 1997](#) and [Hartley 1998](#) studies, some abstracts were rewritten by the original authors of the abstracts. In [Hartley 1997](#), both sorts of rewritten structured abstracts showed significantly better readability scores than the unstructured ones and, in line with other studies, were also significantly longer. Hypothesising that sentences in structured abstracts would contain more positional cues, [Hartley 1997](#) scrambled the sentence order in pairs of structured and unstructured abstracts. Readers made more errors in ordering the sentences from the unstructured version of one of the pairs of abstracts, but this finding did not apply to the second pair and there were mixed results when the study was partially replicated using conference attendees, making the results difficult to interpret. Two of the other studies in [Hartley 1997](#) did not detect significant differences in the readability scores of structured and unstructured abstracts from the *BMJ* and the *British Journal of Psychiatry*, but the numbers of abstracts in these before-and-after studies may have been too small to detect differences. In the final study in [Hartley 1997](#), psychology students found structured

abstracts easier to read than unstructured abstracts (assessed as a mark out of 10). Interestingly, the Flesch Reading Ease scores were very similar for the structured and the unstructured abstracts, suggesting that such scores do not have great face validity when it comes to assessing the readability of abstracts of scientific journal articles. [Hartley 1998](#) studied four psychology journals that introduced a requirement for structured abstracts and they asked authors of accepted papers to revise their abstract in light of the new requirement. Two of the three evaluators of each of 30 pairs of abstracts were undergraduate psychology students and use of students as assessors probably comes closer to using the journals' normal readers than other studies in which assessments have been performed by expert reviewers or journal editors. However [Hartley 1998](#) did not report whether there were any differences (or similarities) between the student evaluators and the third evaluator (the first author of the study). One possible limitation of this study is that all the journals came from one discipline, psychology, and with a concentration from one journal, so the results may have limited generalisability.

There is now quite a substantial body of evidence to indicate that structured abstracts are generally easier to read and contain more information ([Dupuy 2003](#); [Hartley 1997](#); [Hartley 1998](#); [Scherer 1998](#); [Taddio 1994](#); [Trakas 1997](#); [Wong 2005](#)), but are longer than unstructured abstracts ([Comans 1990](#); [Dupuy 2003](#); [Harbourt 1995](#); [Hartley 1997](#); [Hartley 1998](#); [Hartley 2003](#)).

[Booth 1997](#) found that the use of structured abstracts improved searching precision but at the expense of recall, i.e. most relevant papers were identified efficiently, but some relevant papers were missed. The investigators note that this was "a preliminary investigation and therefore carries many of the limitations of such a design. The databases were crude prototypes and the sets of records were too small to sustain detailed statistical examination". There are also some numerical discrepancies between the text of the paper and the tables and we hope to resolve these discrepancies with the study authors, with whom we have made initial contact. Four studies ([Harbourt 1995](#); [Hartley 1996a](#); [Hartley 1996b](#); [Wilczynski 1995](#)) indicate that structured abstracts are easier to search than unstructured ones.

## REFERENCE ACCURACY

The literature contains many surveys and non-comparative observational studies which were excluded from our review because they did not examine the effects of any specified interventions. However they do provide a baseline against which to judge the effectiveness of interventions or indicate deficiencies in the peer-review process which might be remedied or are in need of further research.

The majority of papers identified by our search relate to reference accuracy. We identified 66 papers on this subject which gathered data gathered from over 27,000 references in over 100 different

journals. Although slightly different criteria were employed (for example, some authors included errors of punctuation while others did not) we felt that the methods were sufficiently similar to permit comparisons. Citation error rates in journals ranged from 4% to 67% with a median of 38%. Several studies compared accuracy in a number of journals but did not specify the interventions that might have accounted for any differences observed. However there was some evidence that journals employing in-house checking of references had lower than average error rates, which would support the findings of [Lowry 1985 RA](#); [Hobma 1992 RA](#); and [George 1994 RA](#). Some investigators also assessed the accuracy of quotations to see if cited papers were fairly represented. Quotation error rates per journal ranged from 0% to 50% with a median of 20%, but no authors suggested interventions that might improve the accuracy of citations. One might imagine that one criterion for selecting peer-reviewers would be their knowledge (or even authorship) of the relevant literature, so it would be interesting to measure the accuracy of quotations before and after peer review, to see if it is improved by reviewers' comments.

## OTHER INTERVENTIONS

[Gross 1994](#) studied the effect of including a clinical example on readers' ability to apply information from a paper. While this would not normally fall into the area of technical editing, we included this study because we felt its findings had important implications and this was an area that warranted further research. Although participant numbers were small and there may have been a unit of analysis problem (analysing more than one example per participant), this was quite a well-designed study that indicates that 'enriching' information in particular ways may make research reports more useful.

## WHAT HAS NOT BEEN STUDIED

Apart from [Gross 1994](#), we did not identify any studies comparing different methods of manipulating or presenting data as part of the editing process in peer-reviewed biomedical journals. In a randomised trial of editing versus no editing of radiology reports, [Coakley 2003](#) found that editing significantly improved clarity, brevity, readability and overall impression of quality. It would be interesting to perform similar experiments on the effects of data display formats and presentation on journal readers' perceptions and interpretation of information in research reports. Journals could then recommend the most appropriate format for different types of data.

Several aspects of the technical editing of biomedical journals appear not to have been studied at all. Although journals usually provide detailed instructions to contributors, we found no published work examining the direct effects of these. However there are some studies, mostly using a before-and-after design, which

have investigated changes in the quality of reporting over time. While [Scherer 1998](#) was not able to detect a change in the quality of reporting of clinical trials over time, [Schumm 1999](#) did find a significant improvement in the frequency of reporting of 11 elements of design and analysis from previous baseline surveys ([DerSimonian 1982](#); [Emerson 1984](#)), although the quality of reporting methods remained poor. It is interesting to contrast the effects of 'mass' dissemination of instructions to contributors (with high profile adoption of the CONSORT statement by many journals - [Plint 2006](#)) with the directed form of dissemination of author instructions used in Pitkin's study ([Pitkin 1998](#)). However it is not clear whether the noted improvements were made by authors, by peer reviewers, or by editors and technical editors as part of the editorial process.

Nearly all journals impose a house-style which includes elements of typographic design (such as typeface and page layout) and scientific conventions (such as the use of abbreviations, formats for numbers and the format of references). Again, we found no research about the effects of different styles on legibility or readability in biomedical journals. These aspects may have been researched in other disciplines, but we conclude that, for biomedical journals, the imposition of such styles is not evidence-based, unless there is unpublished research which we have been unable to find. It is conceivable that journals have done in-house research which has not been published. Although we attempted to locate such unpublished studies, our failure to do so is a present limitation of this review. Another apparently unstudied, but widely used, process is proof-reading.

## AUTHORS' CONCLUSIONS

### Implication for systematic reviews and evaluations of healthcare

Surprisingly few studies have evaluated the effects of technical editing rigorously. However there is some evidence that the 'package' of editorial processes used by biomedical journals before publication does improve papers. It is not clear which elements of this 'package' are more or less influential in changing the readability and accessibility of papers.

### Implication for methodological research

Randomized trials comparing discrete parts of the technical editing processes would test their relative contributions to the accessibility and quality of papers (although devising a valid and reliable way to measure quality of papers may be problematic). Such trials could compare different sorts of interventions, or could assess an intervention against no intervention or standard practice. The copy-editing part of the editorial package seems to be the most

urgent component to assess in a randomized trial, although effects of checklists or extra training for copy editors could also be evaluated. The effects of page layout and data display need to be tested with journal articles and journal readers, using qualitative methods.

## ACKNOWLEDGEMENTS

Some of the references for the first version of the review were obtained by the Information Management & Analysis (library) staff of Glaxo Wellcome Research & Development in Greenford.

## REFERENCES

### References to studies included in this review

#### Acea Nebriil 1997 RA *{published data only}*

\* Acea Nebriil B, Gomez Freijoso C. [Errores bibliograficos en la Revista Espanola de Enfermedades Digestivas. Estudio retrospectivo del ano 1995]. *Revista Espanola de Enfermedades Digestivas* 1997;**89**(3):212–4.

#### Aronsky 2005 RA *{published data only}*

\* Aronsky D, Ransom J, Robinson K. Accuracy of references in five biomedical informatics journals. *Journal of the American Medical Informatics Association* 2005;**12**(2):225–8.

#### Asano 1995a RA *{published data only}*

\* Asano M, Mikawa K, Nishina K, Maekawa N, Obara H. The accuracy of references in Anaesthesia. *Anaesthesia* 1995;**50**(12):1080–2.

#### Asano 1995b RA *{published data only}*

\* Asano M, Mikawa K, Nishina K, Maekawa N, Obara H. Improvement of the accuracy of references in the Canadian Journal of Anaesthesia. *Canadian Journal of Anaesthesia* 1995;**42**(5 Pt 1):370–2.

#### Avila 1996 RA *{published data only}*

\* Avila FJ, Pensado A, Esteve C. Errors in bibliographic references in the Revista Espanola de Anestesiologia y Reanimacion: retrospective study of 1994 [Errores en las referencias bibliograficas de la Revista Espanola de Anestesiologia y Reanimacion: estudio retrospectivo del ano 1994]. *Revista Espanola de Anestesiologia Y Reanimacion* 1996;**43**:174–6.

#### Biddle 1996 *{published data only}*

\* Biddle C, Aker J. How does the peer review process influence AANA journal article readability?. *J Am Ass Nurse Anesth* 1996;**64**(1):65–8.

#### Booth 1997 *{published data only}*

\* Booth A, O'Rourke AJ. The value of structured abstracts in information retrieval from MEDLINE. *Health Libraries Review* 1997;**14**:157–66.

#### Browne 2004 RA *{published data only}*

\* Browne RF, Logan PM, Lee MJ, Torreggiani WC. The accuracy of references submitted for publication. *Canadian Association of Radiology Journal* 2004;**55**(3):170–3.

#### Buchan 2005 RA *{published data only}*

\* Buchan JC, Norris J, Kuper K. Accuracy of referencing in the ophthalmic literature. *American Journal of Ophthalmology* 2005;**140**(6):1146–8.

#### Cakir 2003 RA *{published data only}*

\* Cakir V, Ilhan F, Kilicoglu G, Balkan Y, Veziroglu A, Gunal I. [The accuracy of references in Turkish journals in orthopedics and traumatology]. [Turkish]. *Acta Orthopaedica et Traumatologica Turcica* 2003;**37**(4):319–22.

#### Celayir 2003 RA *{published data only}*

\* Celayir AC, Sander S, Celayir S. Accuracy of references in the pediatric surgery journals. *Journal of Pediatric Surgery* 2003;**38**(4):653–4.

#### Comans 1990 *{published data only}*

\* Comans ML, Overbeke AJ. The structured summary: a tool for reader and author [De gestructureerde samenvatting: een hulpmiddel voor lezer en auteur]. *Nederlands Tijdschrift voor Geneeskunde* 1990;**134**(48):2338–40.

#### de Lacey 1985 RA *{published data only}*

\* de Lacey G, Record C, Wade J. How accurate are quotations and references in medical journals?. *BMJ* 1985;**291**:884–6.

#### Doms 1989 RA *{published data only}*

\* Doms CA. A survey of reference accuracy in five national dental journals. *Journal of Dental Research* 1989;**68**(3):442–4.

#### Dupuy 2003 *{published data only}*

\* Dupuy A, Khosrotehanri K, Lebbe C, Rybojad M, Morel P. Quality of abstracts in 3 clinical dermatology journals. *Archives of Dermatology* 2003;**139**:589–93.

#### Eichorn 1987 RA *{published data only}*

\* Eichorn P, Yankauer A. Do authors check their references? A survey of accuracy of references in three public health journals. *American Journal of Public Health* 1987;**77**:1011–2.

- Evans 1990 RA** *{published data only}*  
Evans JT. Quotational and reference accuracy in surgical journals; a continuing peer review problem. Peer review in scientific publishing; papers from the First International Congress in Peer Review in Biomedical Publication. CBE, 1991:75–9.
- Fenton 2000 RA** *{published data only}*  
\* Fenton JE, Brazier H, Desouza A, Hughes JP, McShane DP. The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clinical Otolaryngology* 2000;**25**:40–4.
- Ferreira 2000 RA** *{published data only}*  
\* Ferreira CB, Porto Zocco L. Citations accuracy: a comparative study among Brazilian and international journals on obstetrics and gynecology [abstract]. Eighth International Congress on Medical Librarianship, July 2-5 2000. London, UK, 2000:84.
- Fister 2005** *{published data only}*  
\* Fister K, Marusic A, Hutchings A, Kern J, Marusic M. Editors' impact on improving the accuracy of references: randomized comparison of standard practice, brief reminder, and instructional intervention. International Congress on Peer Review and Biomedical Publication. Chicago, IL, 16–18 September, 2005: 42–3.
- Foreman 1987 RA** *{published data only}*  
\* Foreman MD, Kirchoff KT. Accuracy of references in nursing journals. *Research in Nursing and Health* 1987;**10**:177–83.
- George 1994 RA** *{published data only}*  
\* George PM, Robbins K. Reference accuracy in the dermatologic literature. *Journal of the American Academy of Dermatology* 1994;**31**(1):61–4.
- Goldberg 1993 RA** *{published data only}*  
\* Goldberg R, Newton E, Cameron J, Jacobson R, Bukata WR, Rakab A, et al. Reference accuracy in the emergency medicine literature. *Annals of Emergency Medicine* 1993;**22**(9):1450–4.
- Goodman 1994** *{published data only}*  
\* Goodman SN. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Annals of Internal Medicine* 1994;**121**(1):11–21.
- Goodrich 1977 RA** *{published data only}*  
\* Goodrich JE, Roland CG. Accuracy of published medical reference citations. *Journal of Technical Writing and Communication* 1977;**7**:15–9.
- Gosling 2004 RA** *{published data only}*  
\* Gosling CM, Cameron M, Gibbons PF. Referencing and quotation accuracy in four manual therapy journals. *Manual Therapy* 2004;**9**(1):36–40.
- Gross 1994** *{published data only}*  
\* Gross MT, Sekerak DK, Allen DD. Effect of including a clinical example on the ability of physical therapists to apply information in a technical research report. *Physical Therapy* 1994;**74**:963–8.
- Gupta 2005 RA** *{published data only}*  
\* Gupta P, Yadav M, Mohta A, Choudhury P. References in Indian Pediatrics: Authors need to be accurate. *Indian Pediatrics* 2005;**42**: 140–5.
- Hansen 1994 RA** *{published data only}*  
\* Hansen ME, McIntire DD. Reference citations in radiology: accuracy and appropriateness of use in two major journals. *American Journal of Roentgenology* 1994;**163**(3):719–23.
- Harbourt 1995** *{published data only}*  
\* Harbourt AM, Knecht LS, Humphreys BL. Structured abstracts in MEDLINE(R), 1989-1991. *Bulletin of the Medical Library Association* 1995;**83**(2):190–5.
- Hartley 1996a** *{published data only}*  
\* Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: are structured abstracts more efficient?. *Journal of Information Science* 1996;**22**(5):349–56.
- Hartley 1996b** *{published data only}*  
\* Hartley J, Sydes M, Blurton M. Obtaining information accurately and quickly: are structured abstracts more efficient?. *Journal of Information Science* 1996;**22**(5):349–56.
- Hartley 1996c** *{published data only}*  
\* Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *Journal of Information Science* 1996;**22**(1):27–37.
- Hartley 1997** *{published data only}*  
\* Hartley J, Sydes M. Are structured abstracts easier to read than traditional ones?. *Journal of Research in Reading* 1997;**20**(2): 122–36.
- Hartley 1998** *{published data only}*  
\* Hartley J, Benjamin M. An evaluation of structured abstracts in journals published by the British Psychological Society. *British Journal of Educational Psychology* 1998;**68**:443–56.
- Hartley 2000** *{published data only}*  
\* Hartley J. Could this be easier to read? Tools for evaluating text. In: Hartley J, Branthwaite A editor(s). *The Applied Psychologist*. 2nd Edition. Buckingham: Open University Press, 2000:Chapter 7.
- Hartley 2002** *{published data only}*  
\* Hartley J. Do structured abstracts take more space? And does it matter. *Journal of Information Science* 2002;**28**(5):417–22.
- Hartley 2003** *{published data only}*  
\* Hartley J. Improving the clarity of journal abstracts in psychology. *Science Communication* 2003;**24**(3):366–79.
- Hobma 1992 RA** *{published data only}*  
\* Hobma SO, Overbeke AJPM. Errors in literature references in the Netherlands Tijdschrift voor Geneeskunde [[Fouten in literatuurverwijzingen in het Nederlands Tijdschrift voor Geneeskunde]]. *Nederlands Tijdschrift voor Geneeskunde* 1992;**136**: 637–41.
- Holt 2000 RA** *{published data only}*  
\* Holt S, Siebers R, Suder A, Loan R, Jeffery O. The accuracy of references in Australian and New Zealand medical journals. *New Zealand Medical Journal* 2000;**113**(1119):416–7.
- Jackson 2003 RA** *{published data only}*  
\* Jackson K, Porrino JA, Tan V, Dalusiki A. Reference accuracy in the Journal of Hand Surgery. *Journal of Hand Surgery* 2003;**28A**: 377–80.
- Karlawish 1999** *{published data only}*  
\* Karlawish JHT, Hougham GW, Stocking CB, Sachs GA. What is the quality of the reporting of research ethics in publications of

- nursing home research?. *Journal of the American Geriatrics Society* 1999;**47**(1):76–81.
- Key 1977 RA {published data only}**  
\* Key RD, Roland CG. Reference accuracy in articles accepted for publication in the Archives of Physical Medicine and Rehabilitation. *Archives of Physical Medicine & Rehabilitation* 1977;**58**(3):136–7.
- Khosrotehrani 2002 {published data only}**  
\* Khosrotehrani K, Dupuy A, Lebbe C, Rybojad M, Morel P. Abstract quality assessment of articles from the Annales de Dermatologie [French]. *Annales de Dermatologie et de Venereologie* 2002;**129**(11):1271–5.
- Kolbitsch 1997 RA {published data only}**  
\* Kolbitsch C, Hörmann C, Benzer A. Quotation accuracy in neuroanesthesiology research. *Journal of Neurosurgical Anesthesiology* 1997;**9**:8–10.
- Laccourreye 1999 {published data only}**  
\* Laccourreye O. Evolution of the medical writing of the scientific reports published by the Annales d'Otolaryngologie et de Chirurgie Cervico-Faciale. *Annales d'Oto-Laryngologie et de Chirurgie Cervico-Faciale* 1999;**116**(3):115–25.
- Lawson 1999 RA {published data only}**  
\* Lawson LA, Fosker R. Accuracy of references in psychiatric literature: a survey of three journals. *Psychiatric Bulletin* 1999;**23**(4):221–4.
- Lee 1999 RA {published data only}**  
\* Lee SY, Lee JS. A survey of reference accuracy in two Asian dermatologic journals (the Journal of Dermatology and the Korean Journal of Dermatology). *International Journal of Dermatology* 1999;**38**(5):357–60.
- Lok 2001 RA {published data only}**  
\* Lok CKW, Chan MTV, Martinson I. Risk factors for citation errors in peer-reviewed nursing journals. *Journal of Advanced Nursing* 2001;**34**(2):223–9.
- Lowry 1985 RA {published data only}**  
\* Lowry SR. How accurate are quotations and references in medical journals?. *BMJ* 1985;**291**:1421.
- Lukic 2004 RA {published data only}**  
\* Lukic IK, Lukic A, Gluncic V, Katavic V, Vucenik V, Marusic A. Citation and quotation accuracy in three anatomy journals. *Clinical Anatomy* 2004;**17**(7):534–9.
- McLellan 1992 RA {published data only}**  
\* McLellan MF, Case LD, Barnett MC. Trust, but verify: The accuracy of references in four anesthesia journals. *Anesthesiology* 1992;**77**(1):185–8.
- Mikawa 1996 RA {published data only}**  
\* Mikawa K, Nishina K, Maekawa N, Obara H. Reference accuracy in Intensive Care Medicine [letter]. *Intensive Care Medicine* 1996;**22**(2):176–7.
- Neihouse 1999 RA {published data only}**  
\* Neihouse PF, Priske SC. Quotation accuracy in review articles. *Drug Intelligence and Clinical Pharmacology* 1989;**23**:594–6.
- Ngan Kee 1997a RA {published data only}**  
\* Ngan Kee WD, Roach VJ, Lau TK. How accurate are references in the Australian and New Zealand Journal of Surgery?. *Australian and New Zealand Journal of Surgery* 1997;**67**(7):417–9.
- Ngan Kee 1997b RA {published data only}**  
\* Ngan Kee WD, Roach VJ, Lau TK. The accuracy of references in the Hong Kong Medical Journal. *Hong Kong Medical Journal* 1997;**3**(4):377–80.
- Nishina 1995a RA {published data only}**  
\* Nishina K, Mikawa K, Asano M, Maekawa N, Obara H. Reference accuracy in Critical Care Medicine [letter]. *Critical Care Medicine* 1995;**23**(9):1610–1.
- Nishina 1995b RA {published data only}**  
\* Nishina K, Asano M, Mikawa K, Maekawa N, Obara H. Accuracy of references in Anesthesia & Analgesia does not improve. *Anesthesia and Analgesia* 1995;**80**:641–7.
- Nishina 1995c RA {published data only}**  
\* Nishina K, Asano M, Mikawa K, Maekawa N, Obara H. Improvement of the accuracy of references in Anesthesiology. *Anesthesiology* 1995;**82**:599–600.
- Nishina 1995d RA {published data only}**  
\* Nishina K, Asano M, Mikawa K, Maekawa N, Obara H. The accuracy of references in the Journal of Cardiothoracic and Vascular Anesthesia. *Journal of Cardiothoracic and Vascular Anesthesia* 1995;**9**:622–3.
- Nishina 1995f RA {published data only}**  
\* Nishina K, Mikawa K, Asano M, Maekawa N, Obara H. Reference citation accuracy in the Journal of Anesthesia. *Journal of Anesthesia* 1995;**9**(4):387–9.
- Nishina 2000 RA {published data only}**  
\* Nishina K, Mikawa K, Obara H. Improvement of the accuracy of references in the Journal of Cardiothoracic and Vascular Anesthesia. *Journal of Cardiothoracic and Vascular Anesthesia* 2000;**14**(4):495–6.
- Nuckles 1993 RA {published data only}**  
\* Nuckles DB, Pope NN, Adams JD. A survey of the accuracy of references in 10 dental journals. *Oper-Dent* 1993;**18**(1):28–32.
- O'Connor 2002 RA {published data only}**  
\* O'Connor AE. A review of the accuracy of references in the journal Emergency Medicine. *Emergency Medicine* 2002;**14**(2): 139–41.
- Oermann 2001 RA {published data only}**  
\* Oermann MH, Cummings SL, Wilmes NA. Accuracy of references in four pediatric nursing journals. *Journal of Pediatric Nursing* 2001;**16**(4):263–8.
- Oermann 2002a RA {published data only}**  
\* Oermann MH, Ziolkowski LD. Accuracy of references in three critical care nursing journals. *Journal of Perinatal Nursing* 2002;**17**(2):78–83.
- Oermann 2002b RA {published data only}**  
\* Oermann MH, Mason NM, Wilmes NA. Accuracy of references in general readership nursing journals. *Nurse Educator* 2002;**27**(6): 260–4.

**Oermann 2002c RA {published data only}**

\* Oermann MH, Wilmes NA, Braski P. Reference accuracy in neonatal-maternal nursing literature. *Neonatal Network* 2002;**21**(1):23–6.

**Orlin 1996 RA {published data only}**

\* Orlin W, Pehling J, Pogrel MA. Do authors check their references? A survey of 500 references from the Journal of Oral and Maxillofacial Surgery. *Journal of Oral and Maxillofacial Surgery* 1996;**54**(2):200–2.

**Perez Garcia 2000 RA {published data only}**

\* Perez Garcia A. [Errors in bibliographic references of Nefrologia from 1981 to 1995. A quality control] [Spanish]. *Nefrologia* 2000;**20**(Suppl 6):23–8.

**Pierie 1996 {published data only}**

\* Pierie J-PEN, Walvort HC, Overbeke AJPM. Readers' evaluation of effect of peer review and editing on quality of articles in the Nederlands Tijdschrift voor Geneeskunde. *Lancet* 1996;**348**:1480–3.

**Pieters 2001 RA {published data only}**

\* Pieters G, Ceysens E, De Heyn E. Accuracy and appropriateness of references in the 'Tijdschrift voor Psychiatrie'. *Tijdschrift voor Psychiatrie* 2001;**43**(5):349–53.

**Pitkin 1998 {published data only}**

\* Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions?. *JAMA* 1998;**280**(3):267–8.

**Pitkin 1999 {published data only}**

\* Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999;**281**:1110–1.

**Pitkin 2000 {published data only}**

\* Pitkin RM, Branagan MA, Burmeister LF. Effectiveness of a journal intervention to improve abstract quality. *JAMA* 2000;**283**(4):481.

**Pulida 1995 RA {published data only}**

\* Pulido M, Carles Gonzalez J, Sanz F. Errors in bibliographic references: a retrospective study in Medicina Clinica (1962-1992) [Errores en las referencias bibliograficas: un estudio retrospectivo en Medicina Clinica (1962-1992)]. *Medicina Clinica (Barcelona)* 1995;**104**(5):170–4.

**Putterman 1991 RA {published data only}**

\* Putterman C, Lossos IS. Author, verify your references! or, The accuracy of references in Israeli medical journals. *Israel Journal of Medical Sciences* 1991;**27**(2):109–12.

**Putterman 1992 RA {published data only}**

\* Putterman C. Quotation accuracy: fact or fiction?. *Israel Journal of Medical Sciences* 1992;**28**:465–70.

**Riesenberg 2001 RA {published data only}**

\* Riesenberg LA, Dontineni S. Review of reference inaccuracies. 4th International Congress on Peer Review in Biomedical Publication. Barcelona, September 14–16, 2001.

**Roach 1997 RA {published data only}**

\* Roach VJ, Lau TK, Ngan Kee WD. The quality of citations in major international obstetrics and gynecology journals. *American Journal of Obstetrics and Gynecology* 1997;**177**:973–5.

**Roberts 1994 {published data only}**

\* Roberts JC, Fletcher RH, Fletcher SW. Effects of peer review and editing on the readability of articles published in Annals of Internal Medicine. *JAMA* 1994;**272**:119–21.

**Scherer 1998 {published data only}**

\* Scherer RW, Crawley B. Reporting of randomized clinical trial descriptors and use of structured abstracts. *JAMA* 1998;**280**(3):269–72.

**Schulmeister 1998 RA {published data only}**

\* Schulmeister L. Quotation and reference accuracy of three nursing journals. *Image - the Journal of Nursing Scholarship* 1998;**30**(2):143–6.

**Siebers 1999 RA {published data only}**

\* Siebers R. Accuracy of references in the New Zealand Journal of Medical Laboratory Science. *New Zealand Journal of Medical Laboratory Science* 1999;**53**:46–8.

**Siebers 2000a RA {published data only}**

\* Siebers R. The accuracy of references of three allergy journals. *Journal of Allergy and Clinical Immunology* 2000;**105**:837–8.

**Siebers 2000b RA {published data only}**

\* Siebers R, Holt S. Accuracy of references in five leading medical journals. *Lancet* 2000;**356**:1445.

**Siebers 2001 RA {published data only}**

\* Siebers R. How accurate are references in clinical chemistry?. *Clinical Chemistry* 2001;**47**(3):606–7.

**Siegel 2005 {published data only}**

\* Siegel PZ, Thacker SB, Goodman RA, Gillespie C. Titles of articles in peer-reviewed journals lack essential information: a structured review of contributions to 4 leading medical journals, 1995 and 2001. International Congress on Peer Review and Biomedical Publication. Chicago, IL, 16–18 September 2005:42.

**Silagy 1998 {published data only}**

\* Silagy C, Middleton P, Magarey A, Bastian H. Quality assurance of Cochrane systematic review abstracts; a comparison with abstracts published in Evidence-Based Medicine [abstract]. Sixth International Cochrane Colloquium; 1998 Oct 22-26; Baltimore, MD, USA. 1998:39.

**Sutherland 2000 RA {published data only}**

\* Sutherland AG, Craig N, Maffulli N, Brooksbank A, Moir JS. Accuracy of references in the orthopaedic literature. *Journal of Bone and Joint Surgery (Br)* 2000;**82-B**(1):9–10.

**Taddio 1994 {published data only}**

\* Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *Canadian Medical Association Journal* 1994;**150**(10):1611–5.

**Taylor 1998 RA {published data only}**

\* Taylor MK. The practical effects of errors in reference lists in nursing research journals. *Nursing Research* 1998;**47**:300–3.

**Trakas 1997 {published data only}**

\* Trakas K, Addis A, Kruk D, Buczek Y, Iskedjian M, Einarson TR. Quality assessment of pharmacoeconomic abstracts of original

research articles in selected journals. *Annals of Pharmacotherapy* 1997;**31**:423–8.

**Vargas-Origel 2001 RA** {published data only}

\* Vargas-Origel A, Gomez-Martinez G, Vargas-Nieto MA. The accuracy of references in paediatric journals. *Archives of Disease in Childhood* 2001;**85**(6):497–8.

**Warren 1997 RA** {published data only}

\* Warren KJ, Bhatia N, Teh W, Fleming MG, Lange M. Reference and quotation accuracy in the major and minor infectious diseases journals. Third International Congress on Biomedical Peer Review and Global Communications; Sept 18-20 1997; Prague, Czech Republic. 1997.

**Wilczynski 1995** {published data only}

\* Wilczynski NL, Walker CJ, McKibbin KA, Haynes RB. Preliminary assessment of the effect of more informative (structured) abstracts on citation retrieval from MEDLINE. *MEDINFO* 1995:1457–61.

**Winker 1999** {published data only}

Lantz JC. Unpublished data. November 1998.

**Wong 2005** {published data only}

\* Wong HL, Truong D, Mahamed A, Davidian C, Rana Z, Einarson TR. Quality of structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association: a 10-year follow-up study. *Current Medical Research Opinion* 2005;**21**(4):467–73.

## References to studies excluded from this review

**Bransford 1972** {published data only}

\* Bransford JD, Johnson MK. Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behaviour* 1972;**11**:717–26.

**Broadus 1983 RA** {published data only}

\* Broadus RN. An investigation of the validity of bibliographic citations. *Journal of the American Society for Information Science* 1983;**34**:132–5.

**Cardinal 1995** {published data only}

Cardinal BJ. Readability analysis of health, physical education, recreation and dance journal articles. *Perceptual & Motor Skills* 1995;**80**(1):255–8.

**Charrow 1979** {published data only}

Charrow RP, Charrow VR. Making legal language understandable: a psycholinguistic study of jury instructions. *Colombia Law Review* 1979;**79**:1306–74.

**Coakley 2003** {published data only}

\* Coakley FV, Heinze SB, Shadbolt CL, Schwartz LH, Ginsberg MS, Lefkowitz RA, et al. Routine editing of trainee-generated radiology reports: effect on style quality. *Academic Radiology* 2003;**10**(3):289–94.

**Connatser 1999** {published data only}

Connatser BR. Last rites for readability formulas in technical communication. *J Technical Writing and Communication* 1999;**29**(3):271–87.

**Duffy 1982** {published data only}

Duffy TM, Kabance P. Testing a readable writing approach to text revision. *Journal of Educational Psychology* 1982;**74**:733–48.

**Gould 1984** {published data only}

Gould JD. Doing the same work with hard copy and with cathode ray tube (CRT) computer terminals. *Human Factors* 1984;**26**(3):323–37.

**Hammerschmidt 1992** {published data only}

Hammerschmidt DE, Keane MA. Institutional Review Board (IRB) review lacks impact on the readability of consent forms for research. *American Journal of the Medical Sciences* 1992;**304**(6):348–51.

**Hargens 1990** {published data only}

Hargens LL. Variation in journal peer review systems. Possible causes and consequences. *JAMA* 1990;**263**(10):1348–52.

**Hartley 1998p** {published data only}

\* Hartley J. The role of printouts in editing text. *British Journal of Educational Technology* 1998;**29**(3):277–82.

**Haviland 1972** {published data only}

\* Haviland SE, Clark HH. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behaviour* 1974;**13**:512–21.

**Kauffmann 1991** {published data only}

\* Kauffmann R, Reyes H, Goic A. Editorial analysis of manuscripts sent for publication in the Revista Medica de Chile. *Revista Medica de Chile* 1991;**119**(3):327–33.

**Kronick 1958** {published data only}

\* Kronick DA. Literature citations, a clinico-pathological study, with the presentation of three cases. *Bulletin of the Medical Library Association* 1958;**46**:219–23.

**Macauley 1992** {published data only}

\* Macauley AL, Cullen DJ. Verifying the (we hope) verified: reference checking in the Journal of Clinical Anesthesia. *Journal of Clinical Anesthesia* 1992;**4**:437–8.

**Mohta 2003 RA** {published data only}

\* Mohta A, Mohta M. Accuracy of references in Indian Journal of Surgery. *Indian Journal of Surgery* 2003;**65**(2):156–8.

**Narine 1991** {published data only}

\* Narine L, Yee DS, Einarson TR, Ilersich AL. Quality of abstracts of original research articles in CMAJ in 1989. *Canadian Medical Association Journal* 1991;**144**:449–53.

**Oermann 2005 RA** {published data only}

\* Oermann MH, Wilmes NA. How accurate are references in nursing journals. *Nurse Author and Educator* 2005;**15**(4):1–4.

**Schumm 1999** {published data only}

\* Schumm LP, Fisher JS, Thisted RA, Olak J. Clinical trials in general surgical journals; are methods better reported?. *Surgery* 1999;**125**:41–5.

**Yankauer 1990** {published data only}

\* Yankauer A. The accuracy of medical journal references. *CBE Views* 1990;**13**:38–42.

## References to studies awaiting assessment

**Lee 1995 RA {published data only}**

Lee SY, Lee SJ, Kim YK. A survey of accuracy of reference citations in two Korean dermatological journals. *Annals of Dermatology* 1995;7(3):227–30.

**Nishina 1995e RA {published data only}**

\* Nishina K, Asano M, Mikawa K, Maekawa N, Obara H. The accuracy of reference lists in *Acta Anaesthesiologica Scandinavica*. *Acta Anaesthesiologica Scandinavica* 1995;39(5):577–8.

**Ponce Vargas 1999 RA {published data only}**

Ponce Vargas A, Rodriguez Perez M. References inaccuracies in *Revista Espanola de Reumatologia*: a retrospective study of 1996. *Revista Espanola de Reumatologia* 1999;26(3):82–8.

**Purcell 2001 {published data only}**

\* Purcell GP, Donovan SL, Davidoff F. Changes in manuscripts and quality: the contribution of peer review. Fourth International Congress on Peer Review in Biomedical Publication. Barcelona, September 14–16, 2001.

**Raja 2006 RA {published data only}**

Raja UY, Cooper JG. How accurate are the references in *Emergency Medical Journal*? *Emergency Medical Journal* 2006;23(8):625–6.

**Shulman 1993 RA {published data only}**

Shulman MS, Robillard RJ. Accuracy in reference citations. *Anesthesiology* 1993;78(3):616–7.

**Siebers 2000c RA {published data only}**

Siebers RW. The accuracy of references in the *Australian Journal of Medical Science* 2000;21(2): 16–8.

**Torreggiani 1999 RA {published data only}**

Torreggiani WC, Logan PM, Lee MJ. Accuracy of reference citations in manuscripts submitted for publication. *Irish Journal of Medical Science* 1999;169(Suppl 4):96.

**Additional references****DerSimonian 1982**

DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *New England Journal of Medicine* 1982; 306:1332–7.

**Emerson 1984**

Emerson JD, McPeck B, Mosteller F. Reporting clinical trials in general surgical journals. *Surgery* 1984;95:572–9.

**Godlee 1999**

Godlee F, Jefferson T, editors. *Peer review in health sciences*. London: BMJ Books, 1999.

**Hartley 2000a**

Hartley J. Are structured abstracts more or less accurate than traditional ones? A study in the psychological literature. *Journal of Information Science* 2000;26(4):273–7.

**Jefferson 2007**

Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database of Systematic Reviews* 2007, Issue 2. [DOI: 10.1002/14651858.MR000016.pub3]

**Mago 1999**

Mago R, Crits-Christoph P. Evaluating of reporting and abstracts of clinical trials. *JAMA* 1999;281(1):34–5.

**Overbeke 1999**

Overbeke J. The state of the evidence: what we know and what we don't know about journal peer review. In: Godlee F, Jefferson T editor(s). *Peer review in health science*. London: BMJ Books, 1999: 32–45.

**Plint 2006**

Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia* 2006;185(5):263–7.

**Sackett 2000**

Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. 2nd Edition. Churchill Livingstone, 2000.

\* Indicates the major publication for the study

## CHARACTERISTICS OF STUDIES

### Characteristics of included studies *[ordered by study ID]*

#### Acera Nebril 1997 RA

Methods	Survey of reference accuracy; checked against original sources
Data	100 (91) references randomly selected from Revista Espanola de Enfermedades Digestivas
Comparisons	NA
Outcomes	Citation accuracy
Notes	

#### Aronsky 2005 RA

Methods	Survey of reference accuracy; checked against original sources
Data	656 references in articles from the first 2004 issues of five biomedical informatics journals; Journal of the American Medical Informatics Association, Journal of Biomedical Informatics, International Journal of Medical Informatics, Methods of Information in Medicine, and Artificial Intelligence in Medicine
Comparisons	NA
Outcomes	Citation accuracy
Notes	

#### Asano 1995a RA

Methods	Survey of reference accuracy; checked against original sources
Data	98 references randomly selected from articles published in Anaesthesia in 1990, and 99 references randomly selected from articles published in 1994
Comparisons	NA (but see note)
Outcomes	Citation accuracy
Notes	Editors requested authors to supply photocopies of the first page of each paper cited

**Asano 1995b RA**

Methods	Survey of reference accuracy; checked against original sources; comparative study
Data	94 references randomly selected from articles published in Canadian Journal of Anaesthesia in 1990 and 96 references randomly selected from articles published in 1994
Comparisons	NA, although the 1994 citations were from a period when the editor requested authors to submit first pages of each reference cited
Outcomes	Citation accuracy
Notes	

**Avila 1996 RA**

Methods	Survey of reference accuracy
Data	100 references randomly selected from articles published in REDAR (Revista Espanola de Anestesiologia y Reanimacion)in 1994
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Biddle 1996**

Methods	Before and after study; Single sided t-test; Assessors unaware of which version of each paper (before or after peer review and editorial process)
Data	59 papers (26 case and 33 research reports) published 1992 to 1994 submitted to Journal of the American Association of Nurse Anesthetists
Comparisons	Readability of submitted papers before and after peer review and editorial process
Outcomes	Computer assessment of readability - 59 pairs of papers (Gunning and Flesch scores); Human assessment of readability - 10 pairs of papers (Gunning and Flesch scores)
Notes	“each manuscript is evaluated by a minimum of three advanced practice nurses or physician reviewers, a nurse editor-in-chief, a nonnurse associate editor, a nonnurse publications manager, and the author again at galley proof stage”

**Booth 1997**

Methods	Comparative study; Comparison of three search methods
Data	5 searches in each of two databases, one containing 100 references and one containing 1010 references
Comparisons	Structured search (by using the fields of structured abstracts); unstructured searches; 'gold standard' searches (manual or MEDLINE searches)
Outcomes	Recall (% of 'gold standard' retrieved); Precision (% of references retrieved which were relevant)
Notes	

**Browne 2004 RA**

Methods	Survey of reference accuracy; checked against original sources and indexing tools
Data	259 references from 19 consecutive manuscripts submitted to five radiology journals: American Journal of Roentgenology; Journal of Computer Assisted Tomography; Clinical Radiology; European Radiology; and Canadian Association of Radiologists Journal
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Buchan 2005 RA**

Methods	Survey of reference accuracy; checked against original sources by two independent assessors
Data	200 references (20 randomly selected from each of 10 ophthalmic journals)
Comparisons	NA
Outcomes	Citation and quotation accuracy
Notes	

**Cakir 2003 RA**

Methods	Survey of reference accuracy
Data	182 references randomly selected from four Turkish journals of orthopaedics and traumatology: Acta Orthopaedica et Traumatologica Turcica; Arthroplasty Arthroscopic Surgery; Hacettepe Journal of Orthopaedic Surgery; Journal of Turkish Spinal Surgery
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Celayir 2003 RA**

Methods	Survey of reference accuracy; checked against indexing tools
Data	1506 references from the first issues in 2001 of three paediatric surgery journals: Journal of Pediatric Surgery; Pediatric Surgery International; European Journal of Pediatric Surgery
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Comans 1990**

Methods	Survey of reference accuracy; checked against indexing tools
Data	1506 references from the first issues in 2001 of three paediatric surgery journals: Journal of Pediatric Surgery; Pediatric Surgery International; European Journal of Pediatric Surgery
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**de Lacey 1985 RA**

Methods	Survey of reference accuracy; checked against original sources
---------	---

**de Lacey 1985 RA** (Continued)

Data	300 references from the first issue in 1984 from six medical journals: 50 references were randomly selected from BMJ, Lancet, NEJM, Clinical Radiology, British Journal of Surgery, British Journal of Hospital Medicine
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Doms 1989 RA**

Methods	Survey of reference accuracy; checked against original sources and indexing tools
Data	500 references from the March 1987 issue of five dental journals: 100 references were randomly selected from the Journal of the American Dental Association, Journal of Dentistry for Children, Journal of Dental Research, Journal of Periodontology, Oral Surgery, Oral Medicine, Oral Pathology
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Dupuy 2003**

Methods	Comparative study
Data	49 abstracts from three dermatology journals in 2000 - Archives of Dermatology, British Journal of Dermatology, the Journal of the American Academy of Dermatology
Comparisons	NA
Outcomes	Abstract quality (as measured by a 30-item quality scale divided into 8 categories)
Notes	

**Eichorn 1987 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	150 references from the May 1986 issue of three public health journals: 50 references randomly selected from the American Journal of Public Health, Medical Care, American Journal of Epidemiology

**Eichorn 1987 RA** (Continued)

Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Evans 1990 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	150 references from a single monthly issue in 1987 in three surgical journals: 50 references randomly selected from the American Journal of Surgery ; Surgery; Surgery, Gynecology and Obstetrics
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Fenton 2000 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	200 references randomly selected from the first issues in 1997 of four otolaryngology/head and neck surgery: Laryngoscope; Annals of Otolaryngology, Rhinology and Laryngology; Clinical Otolaryngology; and Journal of Laryngology and Otolaryngology
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Ferreira 2000 RA**

Methods	Survey of reference accuracy; checked against indexing tools
Data	288 references randomly sampled from 1998 issues of obstetrics and gynecology journals: American Journal of Obstetrics and Gynecology, British Journal of Obstetrics and Gynaecology, Revista Brasileira de Ginecologia e

**Ferreira 2000 RA** (Continued)

	Obstetrecia, Femina
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Fister 2005**

Methods	RCT: GENERATION: not reported ALLOCATION: not reported BLINDING: reported to be blinded
Data	75 consecutive manuscripts submitted to a general medical journal
Comparisons	Manuscript was returned to the author with either: 1) standard practice, n=25 manuscripts (prompting authors to acknowledge required changes, with no specific mention of references); 2) brief reminder, n=25 manuscripts (standard practice plus a sentence prompting authors to pay special attention to the accuracy of references; or 3) instructional intervention (standard practice plus a paragraph highlighting the importance of the accuracy of references and a copy of reference citation formats recommended by ICMJE)
Outcomes	Reference quality (complete accuracy, no technical errors, no substantive errors)
Notes	

**Foreman 1987 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	65 references randomly selected from 1983 issues of 10 clinical nursing journals and 47 references randomly selected from 1983 issues of 14 non-clinical nursing journals
Comparisons	NA
Outcomes	citation accuracy
Notes	

**George 1994 RA**

Methods	Survey of reference accuracy; checked against original sources; comparative study
Data	240 references from four dermatology journals in 1990: 60 references randomly selected from the American Academy of Dermatology; Archives of Dermatology; British Journal of Dermatology; Journal of Investigative Dermatology
Comparisons	NA (but see note)
Outcomes	Citation accuracy; quotation accuracy
Notes	two journals monitored reference accuracy and two did not

**Goldberg 1993 RA**

Methods	Survey of reference accuracy; checked against original sources (where possible)
Data	153 references from three emergency medicine journals in 1991: 51 references randomly selected from the American Journal of Emergency Medicine; Annals of Emergency Medicine; Journal of Emergency Medicine
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Goodman 1994**

Methods	Before and after study; Before and after versions were randomly assigned to assessors (with assessors not aware of which version they received) ; Linear regression used to assess the effect of initial quality on the before-after change
Data	111 consecutive research papers accepted for publication in the Annals of Internal Medicine between March 1992 and March 1993
Comparisons	Manuscript quality before and after the editorial process

**Goodman 1994** (Continued)

Outcomes	Manuscript quality, using an assessment tool of 34 items with 44 assessors. Each item was rated on a five-point scale. Percentage of items scoring 3 points or higher; Average of all score components; Dichotomised item scores (0=2 or less, 1=3 or more);
Notes	Quality was defined as 'whether the authors have described their research in enough detail and with sufficient clarity so a reader could make an independent judgment about the strengths and weaknesses of their data and conclusions'

**Goodrich 1977 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	2195 references from the first article of 10 consecutive issues of 10 North American biomedical journals. Three journals were major general medical journals (Annals of Internal Medicine; JAMA; and NEJM) and seven journals represented some of the chief medical specialty areas (American Journal of Psychiatry; American Journal of Public Health; Anesthesiology; Journal of Bone and Joint Surgery; Journal of Medical Education; Pediatrics; Surgery, Gynecology and Obstetrics).
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Gosling 2004 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	320 references randomly sampled from four manual therapy journals (80 references from each of the Journal of Bodywork and Movement Therapies; Journal of Manipulative and Physiological Therapeutics; Journal of Osteopathic Medicine; and Manual Therapy).
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Gross 1994**

Methods	RCT; 69 therapists randomly assigned to two groups; 41 (60%) responded (20 in group 1 and 21 in group 2)
---------	---

**Gross 1994** (Continued)

Data	35 therapists assigned to group 1 (article with application example) and 34 therapists assigned to group 2 (article without application example)
Comparisons	Report submitted to Physical Therapy with and without a section on application examples (of how to calculate values predicting muscle performance)
Outcomes	Selection of appropriate model; Problem solving (ability to derive torque values); Use of article in practice
Notes	Unit of analysis problem? - results presented as observations (2 observations per therapist)

**Gupta 2005 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	Original articles from all 12 issues of Indian Pediatrics from 2002
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Hansen 1994 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from the June 1993 issues of two radiology journals: 47 references from the American Journal of Roentgenology; and 48 references from Radiology were checked
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Harbourt 1995**

Methods	Study of characteristics of structured abstracts in MEDLINE
Data	All 924, 748 MEDLINE records indexed from March 1989 to December 1991
Comparisons	NA
Outcomes	Access points for indexing; Length of abstract
Notes	

**Hartley 1996a**

Methods	Comparative study
Data	52 readers and 8 structured and 8 unstructured abstracts
Comparisons	NA
Outcomes	Reading time; comprehension errors
Notes	

**Hartley 1996b**

Methods	Comparative study
Data	56 readers asked to find particular abstracts from a database
Comparisons	NA
Outcomes	Reading time; comprehension errors
Notes	

**Hartley 1996c**

Methods	Comparative study (prospective comparison of different versions of an abstract)
Data	Three substudies - part 1 assessed the preferences of four groups of readers with 32 participants in each group - part 2 assessed the preferences of 60 readers (first enquiry) and two groups each of 41 readers (second enquiry) - part 3 assessed the preferences of 75 readers
Comparisons	Comparison of multiple versions of the layout and typography for an abstract

**Hartley 1996c** (*Continued*)

Outcomes	Preferences for various presentations of an abstract
Notes	

**Hartley 1997**

Methods	Set of 8 comparative studies (2 retrospective before-and-after studies and 6 prospective comparative studies)
Data	Abstracts from BMJ, British Journal of Psychiatry; samples of abstracts from previous studies
Comparisons	Unstructured (traditional) abstracts versus structured abstracts (rewritten by investigators or original authors)
Outcomes	Flesch reading ease; Gunning Fog index; length (in words); ordering errors; ease of reading (out of ten points)
Notes	The eight studies were treated as one overall study for the purposes of this review

**Hartley 1998**

Methods	Before-and-after study across four journals
Data	30 pairs of abstracts (unstructured and structured) from four psychology journals
Comparisons	Unstructured (traditional) abstract written when the paper was submitted compared with a structured abstract for the same paper, written when the paper was revised
Outcomes	Flesch reading ease; Gunning Fog index; length (in words); evaluation score; time taken to evaluate; qualitative assessment of authors' views
Notes	

**Hartley 2000**

Methods	Comparative study
Data	30 pairs of abstracts (unstructured and structured) from journals published by the British Psychological Society

**Hartley 2000** (Continued)

Comparisons	Unstructured (traditional) abstract written when the paper was submitted compared with a structured abstract for the same paper, written by each original author
Outcomes	Accuracy (inconsistencies between abstract and text; data in abstract not in text; unjustified conclusions)
Notes	Few inaccuracies in either set of abstracts

**Hartley 2002**

Methods	Comparative study
Data	10 or more articles in each of 15 journals
Comparisons	Increasing length of unstructured abstracts; or decreasing length of structured abstracts
Outcomes	Change in pagination for article
Notes	

**Hartley 2003**

Methods	Comparative study
Data	24 unstructured abstracts from the Journal of Educational Psychology
Comparisons	Unstructured abstracts rewritten as structured abstracts
Outcomes	Abstract length; sentence length; percentage of passives; Flesch reading score; use of longer words; use of common words; use of present tense; information checklist; clarity ratings
Notes	

**Hobma 1992 RA**

Methods	Survey of reference accuracy; checked against Index Medicus for reference accuracy and original sources for quotation accuracy; comparative study
Data	100 references randomly selected from articles published in Nederlands Tijdschrift voor Geneeskunde in volume 135 (1991) and 100 references submitted for publication in a four-week period in 1991. The references in the published articles were sorted into two groups of 50 references, with one group consisting of 14 original articles with a maximum of 12 references and the other group consisting of 10 original articles with 25 or more references.
Comparisons	Submitted versus published articles
Outcomes	Citation accuracy
Notes	

**Holt 2000 RA**

Methods	Survey of reference accuracy; if found, references were checked against the PubMed database (U.S. National Library of Medicine)
Data	References from articles in the August 1999 issues of three New Zealand and Australian medical journals: 188 references (out of a total of 268) were checked for the New Zealand Medical Journal, 430 references (out of 551) for the Medical Journal of Australia, and 404 references (out of 470) for the Australian and New Zealand Journal of Medicine
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Jackson 2003 RA**

Methods	Survey of reference accuracy; checked against original sources Comparison between 1985 and 1995 (in 1995 the Journal of Hand Surgery began requesting authors to supply a copy of the first page of each journal article of book cited)
Data	100 references randomly selected from each of the 1985 and 1995 Journal of Hand Surgery
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Karlawish 1999**

Methods	Quality assessment;
Data	45 publications of research involving nursing home residents, in 4 journals
Comparisons	Comparison of instructions to authors (in 4 journals) regarding ethics with quality of reporting research ethics
Outcomes	Measurement of four aspects of the quality of research ethics; justification of use of nursing home residents; informed consent obtained or waived; IRB review; nursing home committee review
Notes	

**Key 1977 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	1867 references from the March 1975 and March 1976 issues of the Archives of Physical Medicine and Rehabilitation
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Khosrotehrani 2002**

Methods	Before and after study
Data	Assessed abstract quality in Annales de Dermatologie before and after the introduction of structured abstracts in 1993 (total of 43 abstracts)
Comparisons	NA (3 time periods - 1991-2; 1996; 2000)
Outcomes	Abstract quality score (based on Narine)
Notes	

**Kolbitsch 1997 RA**

Methods	Survey of reference accuracy
Data	Used Science Citation Index to track 32 subsequent references to a single article published in 1973. These references were in articles published between 1974 and 1995 in six anaesthesia journals - Acta Anaesthesia Scandinavica (4 articles); Anaesthesia (2 articles); Anesthesia Analgesia (6 articles); Anesthesiology (6 articles); British Journal of

**Kolbitsch 1997 RA** *(Continued)*

	Anaesthesia (12 articles); Journal of Neurosurgery and Anesthesiology (2 articles).
Comparisons	NA
Outcomes	Quotation accuracy
Notes	

**Laccourreye 1999**

Methods	Comparative study over three time points
Data	98 scientific reports published in the Annales d'Otolaryngologie et de Chirurgie Cervico-faciale in 1977, 1987 and 1997
Comparisons	Introduction of stricter editorial policies
Outcomes	Standard of medical writing measured qualitatively and quantitatively
Notes	

**Lawson 1999 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	50 references randomly selected from 1997 issues each of three psychiatric journals (Psychiatric Bulletin; British Journal of Psychiatry; American Journal of Psychiatry)
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Lee 1999 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	200 references from 1993 issues of two dermatology journals; 100 references from the Journal of Dermatology; and 100 references from the Korean Journal of Dermatology
Comparisons	NA

**Lee 1999 RA** (*Continued*)

Outcomes	Citation accuracy; quotation accuracy
Notes	

**Lok 2001 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	550 references randomly selected from 1998 issues of 11 nursing journals
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Lowry 1985 RA**

Methods	Survey of reference accuracy; comparative study
Data	All direct quotations and references from 28 letters to the editor of the BMJ received in the week beginning 7 May 1984 and 61 letters published in the BMJ in the week beginning 7 May 1984
Comparisons	Comparison of correspondence that was received and correspondence that was published, in the same time period
Outcomes	Citation accuracy; quotation accuracy
Notes	

**Lukic 2004 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	199 randomly selected references from 2001 issues three anatomy journals: Annals of Anatomy; Clinical Anatomy; Surgical and Radiologic Anatomy
Comparisons	NA
Outcomes	Citation accuracy; quotation accuracy
Notes	

**McLellan 1992 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	400 references (100 per journal) randomly selected from 22,478 references from all articles in the 1988 issues of four anaesthesia journals; after excluding references to nonjournal articles, a total of 348 references were checked, consisting of 87 references from Anesthesiology, 86 from Anesthesia and Analgesia, 91 from the British Journal of Anaesthesia, and 84 from the Canadian Journal of Anaesthesia.
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Mikawa 1996 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from all 2,427 references in articles published in the 1993 issues of Intensive Care Medicine, with 94 references checked
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Neihouse 1999 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from 54 review articles (citing a total of 3952 references) of drug therapy published from January to December 1987 in four drug journals: 40 references from Clinical Pharmacy, 25 references from Drug Intelligence Clinical Pharmacology, 23 references from Drugs and 12 references from Pharmaco-therapy
Comparisons	NA
Outcomes	Quotation accuracy
Notes	

**Ngan Kee 1997a RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from all 4092 references in the 1995 issues of the Australian and New Zealand Journal of Surgery; 90 references were checked
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Ngan Kee 1997b RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from each of the 1995 and 1996 volumes of the Hong Kong Medical Journal
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Nishina 1995a RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from 8771 references in the 1993 issues of Critical Care Medicine; 96 references were checked
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Nishina 1995b RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from 5343 references in the 1993 issues of Anesthesia and Analgesia; and 100 references randomly selected from 5737 references in the 1994 issues; 96 references were checked for 1990, and 97 references were checked for 1994

**Nishina 1995b RA** (Continued)

Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Nishina 1995c RA**

Methods	Survey of reference accuracy; checked against original sources; comparative study
Data	100 references randomly selected from 11,060 references in the 1990 issues of Anesthesiology, and 100 references randomly selected from 5523 references in the 1994 issues; 96 references were checked for 1990 and 97 references for 1994
Comparisons	NA (but see note)
Outcomes	Citation accuracy
Notes	Editors requested authors to check references against original sources

**Nishina 1995d RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from 2465 references in the 1990 issues of the Journal of Cardiothoracic and Vascular Anesthesia, and 100 references randomly selected from 2079 references in the 1993 issues; 98 references were checked for 1990 and 97 for 1993
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Nishina 1995f RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from 1987 and 1994 issues of the Journal of Anaesthesia
Comparisons	NA

**Nishina 1995f RA** (Continued)

Outcomes	Citation accuracy
Notes	

**Nishina 2000 RA**

Methods	Survey of reference accuracy; checked against original sources; comparative study
Data	100 references randomly selected from 3,618 references in the 1998 issues of the Journal of Cardiothoracic and Vascular Anesthesia, and 100 references randomly selected from 3,433 references in the 1999 issues; 98 references were checked for 1998 and 97 references for 1999
Comparisons	NA (but see note)
Outcomes	Citation accuracy
Notes	Editors requested authors to check references against original sources

**Nuckles 1993 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	30 references randomly selected from the January 1991 issue of 10 dental journals (total of 300 references)
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**O'Connor 2002 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references selected from volume 12 of Emergency Medicine
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Oermann 2001 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	190 references selected from four pediatric nursing journals
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Oermann 2002a RA**

Methods	Survey of reference accuracy; checked against original sources
Data	10% of references randomly selected from Journal of Perinaesthesia Nursing; American Journal of Critical Care; Critical Care Nurse (total of 244 references)
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Oermann 2002b RA**

Methods	Survey of reference accuracy; checked against original sources
Data	130 references selected from three general readership nursing journals - American Journal of Nursing; Nursing Outlook, RN
Comparisons	NA (but see note)
Outcomes	Citation accuracy
Notes	Some journals had a librarian checking the references in submitted manuscripts

**Oermann 2002c RA**

Methods	Survey of reference accuracy; checked against original sources
Data	221 references randomly selected from three nursing journals - Neonatal Network: The Journal of Neonatal Nursing; Journal of Obstetric, Gynecologic and Neonatal Nursing; American Journal of Maternal/Child Nursing

**Oermann 2002c RA** (Continued)

Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Orlin 1996 RA**

Methods	Survey of reference accuracy; checked against indexes or original sources
Data	500 references randomly selected from 1992 issues of the Journal of Oral and Maxillofacial Surgery
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Perez Garcia 2000 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	5 references from each of 87 articles randomly selected from 1981 to 1995 issues of Nefrologia
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Pierie 1996**

Methods	Comparative study (survey); Journal readers each evaluated the quality of 3 versions (submitted, accepted and published) of two articles. The 3 versions were packed in random order and were blinded for authors, research institute and type of version. Differences in scores measured by McNemar's test ( $p < 0.05$ )
Data	100 volunteer readers of Nederlands Tijdschrift voor Geneeskunde (25 medical students, 25 medical graduates, 25 general practitioners, 25 specialists)
Comparisons	Quality of submitted article compared to quality of accepted article; Quality of accepted article with quality of published article
Outcomes	25 questions, each with a five point scale; maximum of 4 evaluators x 50 articles = 200 evaluations per question

**Pierie 1996** (Continued)

Notes	Percentages only given (based on 128 to 196 evaluations per question), but details of numbers of evaluations for or each question not provided A minimum sample size of 34 articles was implied from assuming that an average-scoring item had a 50% chance of being acceptable prior to, and a 90% chance after, peer review and editing.
-------	---

**Pieters 2001 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from three issues of Tijdschrift voor Psychiatrie
Comparisons	NA
Outcomes	Citation accuracy; Quotation accuracy
Notes	

**Pitkin 1998**

Methods	RCT; GENERATION computer generated list of random numbers ALLOCATION Clerk assigning codes was not involved in the study and codes were not broken until study completion BLINDING Outcome assessors masked with respect to assignment to intervention or control group
Data	250 manuscripts for Obstetrics and Gynecology, reporting original research returned to authors with an invitation to revise, between August 12 1994 to December 5, 1995. Final numbers available for analysis = 203; 89 instructed 114 uninstructed
Comparisons	Inclusion of printed instructions for authors preparing abstracts versus no inclusion
Outcomes	Proportion of abstracts containing 1 or more of the following defects; inconsistency in between abstract and text, tables or figures, data in abstract but not body, conclusions not justified by information in the abstract
Notes	

**Pitkin 1999**

Methods	Comparative study
Data	264 articles (44 from each of 5 journals): Annals of Internal Medicine, BMJ, JAMA, Lancet, NEJM
Comparisons	Differences between journals
Outcomes	Deficiencies in abstracts
Notes	Sample size calculated on assumption of a 10-40% range of deficient abstracts across the journals

**Pitkin 2000**

Methods	Before and after study; date, volume and page numbers of articles were masked and were numbered according to a computer-generated set of random numbers
Data	100 articles; first 50 original contributions in JAMA 1998;278 and last 50 original contributions in JAMA 1998;280
Comparisons	Quality improvement initiative in JAMA;
Outcomes	Overall deficiencies in abstracts; non-trivial deficiencies; data inconsistent between abstract and text; both above 2 deficiencies; unjustified conclusions
Notes	k = 0.89 for agreement between the 2 evaluators

**Pulida 1995 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	5 references systematically sampled from original articles in 1962 to 1992 volumes of Medicina Clinica
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Putterman 1991 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	The first reference was selected from 384 articles in the 1986 issues of two general medical journals published in Israel; Harefuah and the Israel Journal of Medical Sciences. These 384 references represented 6.2% of the references in 1986 issues of Harefuah (209/3,345) and 6.2% of the references in 1986 issues of the Israel Medical Journal (175/2,814)
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Putterman 1992 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	120 references randomly selected from articles published in the January 1990 issue of the two general medical journals published in Israel; Harefuah and the Israel Journal of Medical Sciences (60 references per journal)
Comparisons	NA
Outcomes	Quotation accuracy
Notes	

**Riesenberg 2001 RA**

Methods	Review of studies of reference accuracy
Data	30 studies of reference accuracy from 1979 to 2000
Comparisons	NA
Outcomes	Citation accuracy; Quotation accuracy
Notes	

**Roach 1997 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	150 references (50 per journal) were randomly selected from the 1995 issues of three obstetrics and gynaecology journals: 45 references from the Australian and New Zealand Journal of Obstetrics and Gynaecology, 46 references from the American Journal of Obstetrics and Gynecology and 42 references from the British Journal of Obstetrics and Gynaecology were checked
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Roberts 1994**

Methods	Before and after study
Data	101 consecutive manuscripts reporting original research, for Annals of Internal Medicine between March 1 and November 30, 1992
Comparisons	Peer review and editorial processes of the Annals of Internal Medicine (at least one editor-in-chief, a deputy editor, at least one associate editor, at least two reviewers, a statistician and at least one copy editor)
Outcomes	ABSTRACTS Gunning fog index; Flesch reading ease score; syllables/word; words/sentence; total words (median, range) MANUSCRIPTS Gunning fog index; Flesch reading ease score; syllables/word; words/sentence; total words (median, range)
Notes	

**Scherer 1998**

Methods	Before-and-after study; probably not masked
Data	125 reports of RCTs in 3 ophthalmology journals; 77 reports for the structured/unstructured comparison and 48 from the 1991/2 with 1993/4 comparison
Comparisons	Unstructured versus structured abstracts (Archives of Ophthalmology and Ophthalmology); change over time in unstructured abstracts (American Journal of Ophthalmology)
Outcomes	Quality of reporting in the abstracts and texts of RCTs: Number of relevant CONSORT criteria (out of a total of 9) included in the abstract; Number of relevant criteria (out of a total of 56) included in the text

**Scherer 1998** (Continued)

Notes	
-------	--

**Schulmeister 1998 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	180 references randomly selected from the July 1995 to June 1996 issues in three nursing journals; Image: Journal of Nursing Scholarship, Nursing Management, RN
Comparisons	NA
Outcomes	Citation accuracy; Quotation accuracy
Notes	

**Siebers 1999 RA**

Methods	Survey of reference accuracy; checked against bibliographic databases or original sources
Data	99 references from all leading, review and original articles published in the New Zealand Journal of Medical Laboratory Science from May 1998 to April 1999
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Siebers 2000a RA**

Methods	Survey of reference accuracy; checked against bibliographic databases or original sources
Data	1,787 references from articles published in the April 1999 issues of three allergy journals (788 references from the Journal of Allergy and Clinical Immunology, 589 references from Clinical and Experimental Allergy and 410 references from Allergy)
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Siebers 2000b RA**

Methods	Survey of reference accuracy; References which appeared in MEDLINE were checked against MEDLINE, with mismatches checked against other bibliographic databases or original sources
Data	1,557 references from the first issue in March 1999 of five leading general medical journals (395 references from the New England Journal of Medicine; 280 from the Annals of Internal Medicine; 213 from the BMJ, 317 from JAMA and 352 from the Lancet)
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Siebers 2001 RA**

Methods	Survey of reference accuracy; checked against MEDLINE
Data	All 1063 references from the December 1999 issue of Clinical Chemistry
Comparisons	NA
Outcomes	Citation accuracy
Notes	

**Siegel 2005**

Methods	Comparative study between journals of the information content of journal titles
Data	Titles of articles from BMJ, JAMA, Lancet and the New England Journal of Medicine in 1995 and 2001
Comparisons	NA
Outcomes	Information content (topic only, methods, results, conclusions, data set) of titles
Notes	

**Silagy 1998**

Methods	Before and after study
Data	15 abstracts of Cochrane reviews from 1995 to March 1998 professionally edited for the journal Evidence-Based Medicine
Comparisons	NA

**Silagy 1998** (Continued)

Outcomes	Number of words; Flesch Reading Ease index; Change in quantity of information; Change in meaning of information
Notes	

**Sutherland 2000 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly sampled from 1985 to 1995 issues of the Journal of Bone and Joint Surgery [British], the Journal of Bone and Joint Surgery [American], and Clinical Orthopaedics and Related Research; and from 1985 to 1994 issues of Acta Orthopaedica Scandinavica
Comparisons	NA
Outcomes	Citation accuracy
Notes	Reports total number of errors in references, not how many references contain errors

**Taddio 1994**

Methods	Before-and-after study; partially masked
Data	Abstracts from BMJ, CMAJ, JAMA
Comparisons	Unstructured versus structured abstracts
Outcomes	Set of quality criteria
Notes	

**Taylor 1998 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	Stratified random sample of 10% of the total 2,623 references in all articles published in the second half of 1994 in three nursing journals; 262 references (87 references from Image: Journal of Nursing Scholarship; 92 from Nursing Research; and 83 from Western Journal of Nursing Research)
Comparisons	NA

**Taylor 1998 RA** (Continued)

Outcomes	Citation accuracy
Notes	

**Trakas 1997**

Methods	Before-and-after study; partially masked
Data	51 pharmaco-economics articles in 10 journals
Comparisons	Unstructured versus structured abstracts
Outcomes	Set of quality criteria
Notes	

**VargasOrigel 2001 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	100 references randomly selected from each of four paediatric journals: 1999 issues of Acta Paediatrica, Archives of Disease in Childhood, Journal of Pediatrics, Pediatrics
Comparisons	NA
Outcomes	Citation error
Notes	

**Warren 1997 RA**

Methods	Survey of reference accuracy; checked against original sources
Data	240 references selected from 4 major infectious diseases journals (Clinical Infectious Diseases, Journal of Infectious Diseases, Scandinavian Journal of Infectious Diseases and Pediatric Infectious Disease Journal) and 142 references selected from 3 minor infectious disease journals (Infections in Medicine, Infections in Urology, and the AIDS Reader) and 3 minor specialty journals (Complications in Surgery, Pediatric Annals and Complications in Orthopedics)
Comparisons	NA
Outcomes	Citation accuracy; Quotation accuracy

**Warren 1997 RA** (Continued)

Notes	
-------	--

**Wilczynski 1995**

Methods	Comparison of searching performance (citation retrieval) between structured and unstructured abstracts - over time and between journals
Data	All articles in 10 internal and general medicine articles in 1986 and 1991
Comparisons	NA
Outcomes	Sensitivity, specificity and precision of search terms (judged against the 'gold standard' of a manual review of articles)
Notes	

**Winker 1999**

Methods	Before-and-after study
Data	Articles in JAMA
Comparisons	Quality improvement initiative in JAMA
Outcomes	Discrepancies between text and abstract
Notes	

**Wong 2005**

Methods	Before and after study of quality of abstracts; raters were blinded
Data	Abstracts from 1991/2 and 2001/2 issues of BMJ, Canadian Medical Association Journal and the Journal of the American Medical Association (also compared with 1988/9 unstructured abstracts)
Comparisons	NA
Outcomes	Abstract quality (as used by Taddio 1994)
Notes	

ICMJE = International Committee of Medical Journal Editors IRB = Institutional Review Board  
NA = Not Applicable NEJM = New England Journal of Medicine RCT = Randomised Controlled Trial

### Characteristics of excluded studies *[ordered by study ID]*

Bransford 1972	Not biomedical
Broadus 1983 RA	No results could be extracted
Cardinal 1995	Baseline survey with no comparisons
Charrow 1979	Not biomedical
Coakley 2003	Radiology reports not journal articles
Connatser 1999	Not biomedical
Duffy 1982	Not biomedical
Gould 1984	Not biomedical
Hammerschmidt 1992	Patient information, not biomedical journal articles
Hargens 1990	Comparison of acceptance rates and times to acceptance in three journals, but these were not biomedical journals
Hartley 1998p	No comparisons, not biomedical
Haviland 1972	Not biomedical
Kauffmann 1991	No comparison - baseline survey
Kronick 1958	Descriptive; no data could be extracted
Macauley 1992	No data; comment on a study
Mohta 2003 RA	Reference accuracy study; counted punctuation errors so unable to be compared with other reference accuracy studies
Narine 1991	No comparison - baseline survey
Oermann 2005 RA	Review of four surveys of reference accuracy (all included in this review)
Schumm 1999	Reporting quality, not technical editing
Yankauer 1990	No original data; comment

## DATA AND ANALYSES

### Comparison 1. Technical editing

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 study results			Other data	No numeric data
1.1 Peer review and editing reports			Other data	No numeric data
1.2 Providing instructions to authors			Other data	No numeric data
1.3 Providing instructions to readers			Other data	No numeric data
1.4 Structuring abstracts			Other data	No numeric data

### Comparison 2. Citation and quotation accuracy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Error rates (proportion of incorrect references)			Other data	No numeric data

#### Analysis 1.1. Comparison 1 Technical editing, Outcome 1 study results.

##### study results

Peer review and editing reports	
Biddle 1996	<p>All of the following results (reported as means and standard deviations) showed significant changes (<math>p &lt; 0.01</math>) with a one-tailed t test. (Increased ease of reading is indicated by lower Gunning Fog scores but by higher Flesch Reading Ease scores.)</p> <p>Computer analysis</p> <p>Gunning Fog (26 case reports): before editing 18.20 (3.11), after editing 15.98 (3.85)</p> <p>Gunning Fog (33 research reports): before editing 19.36 (2.94), after editing 14.90 (2.63)</p> <p>Flesch Reading Ease (26 case reports): before editing 27.14 (8.60), after editing 33.79 (5.64)</p> <p>Flesch Reading Ease (33 research reports): before editing 24.60 (9.34), after editing 32.45 (6.44)</p> <p>Human analysis</p> <p>Gunning Fog (10 research reports): before editing 18.23 (6.47), after editing 15.85 (7.34)</p> <p>Flesch Reading Ease (10 research reports): before editing 26.92 (5.16), after editing 35.78 (11.37)</p> <p>Word length (26 case reports): before editing 2793 (973), after editing 2371 (840)</p>

**study results** (Continued)

	Word length (33 research reports): before editing 4842 (1225), after editing 3609 (1043)
George 1994 RA	Error rate of 35% in monitored journals and 48% in journals that did not monitor citations, p=0.066
Goodman 1994	<p>The percentage of manuscripts scoring more than 3 on a 5-point scale rose by 7.3% (95% CI 3.3 to 11.3) from a baseline of 75% (before peer review and editing). The average item score improved by 0.23 points (95% CI 0.07 to 0.39) from a baseline score of 3.5 (out of a possible 5). A subjective 10-point global score of quality did not show a statistically discernible change, increasing by 0.29 units (95% CI -0.25 to 0.83), p = 0.3, after peer review and editing. Lower quality manuscripts showed more improvement after peer review and editing than did higher quality manuscripts.</p> <p>The largest changes in the 34-item instrument after peer review and editing were seen in:</p> <ul style="list-style-type: none"> <li>- Discussion of study limitations (47% to 65%, p&lt;0.001)</li> <li>- Acknowledgment and justification of generalisations (58% to 79%, p&lt;0.001)</li> <li>- Appropriateness of the strength or tone of the conclusions (71% to 85%, p=0.01)</li> <li>- Use of confidence intervals (65% to 81%, p&lt;0.001).</li> </ul>
Hobma 1992 RA	Submitted articles contained 70 (70/100) citation errors compared to 31/100 (31%) in published articles
Jackson 2003 RA	After requiring a copy of the first page of each reference, the error rate fell from 30% (30/100) in 1985 to 11% in 1995
Laccourreye 1999	Median of 1.2 errors per page in 1977, 2.2 in 1987 and 2.5 in 1997. The percentage of articles following the IMRAD (Introduction, Methods, Results And Discussion) structure also showed a statistically significant increase over time, with 100% of the 1997 reports (n=14) following IMRAD. Stricter editorial policies were introduced by the journal in 1990 and Uniform Requirements for Publishing (ICMJE 1991) were also released in 1990.
Lowry 1985 RA	<p>Quotation error in correspondence received 7/25 (28%); in correspondence published 7/61 (12%)</p> <p>Citation error in correspondence received 7% (5/67); in correspondence published 3% (7/248)</p> <p>Overall 69% of references in letters received were completely accurate compared to 92% in published letters</p>
Pierie 1996	<p>The 14 questions showing significant improvement dealt with: Introduction (background); Methods (setting, definitions); Results (outcome, statistics, understandability, numerical data); Discussion (significance, other 'proof', limitations); General (abstract, length, general medical value, overall).</p> <p>Questions not showing a statistically significant difference dealt with: Introduction (objective); Methods (inclusion, distinction groups, design); Results (description, tables and figures); Discussion (conclusions, importance of conclusions); General (title).</p> <p>The 11 editing questions showing significant improvement dealt with: Readability (readability, style); Methods (setting, design, measurement technique); Results (presentation, tables and graphs, numerical data); Discussion (conclusion); General (title, references)</p> <p>Questions not showing a statistically significant difference dealt with: Readability (terms, organisation); Methods (time); Results (differences); General (abstract).</p>

**study results** (Continued)

Pitkin 1999	Journal A : 8 deficient abstracts out of 44 (18%, 95% CI 6 to 30) Journal B: 19 deficient abstracts out of 44 (43%, 95% CI 29 to 58) Journal C: 13 deficient abstracts out of 44 (30%, 95% CI 16 to 43) Journal D: 20 deficient abstracts out of 44 (45%, 95% CI 30 to 59) Journal E: 14 deficient abstracts out of 44 (32%, 95% CI 18 to 45) Journal F: 30 deficient abstracts out of 44 (68%, 95% CI 54 to 82) The chi-square test shows a statistically significant difference between journals (chi-square, with 5 degrees of freedom = 31.3, p<0.001).
Pitkin 2000	All types of non-trivial deficiencies, except unjustified conclusions, showed decreases: Data inconsistent between abstract and text dropped from 8/50 to 5/50 Data present in abstract but not present in text dropped from 9/50 to 1/50 Abstracts containing both the above deficiencies dropped from 8/50 to 3/50 Unjustified conclusions were present in one of the 50 abstracts both before and after the quality improvement initiative.
Roberts 1994	The Gunning Fog score for the main text improved from 17.16 (SD 1.55) before editing to 16.85 (1.42) after editing (p=0.0005), a lower score indicates greater readability but both scores remained in the 'very difficult' category). The Flesch Reading Ease score was 28.19 (7.89) before editing, improving to 29.11 (7.73) afterwards (p = 0.03). A higher score represents improved readability, although the score after editing just moved from the 'very difficult' to 'difficult' category). The number of words per sentence also dropped significantly after peer review and editing, but there was a small overall increase in the length of both the main text and the abstract.
Siegel 2005	Only one of four major journals (BMJ) showed a significant increase in the number of journal titles that contained information about study methods (increase from 49% (n=133) in 1995 to 96% (n=112) in 2001, p < 0.001).
Silagy 1998	15 abstracts of Cochrane reviews (CR) edited by the journal Evidence-Based Medicine (EBM) were shorter than the original (330 EBM, 378 CR); and more readable (mean Flesch Reading Ease score of 35.9 EBM, 33.6 CR).
Winker 1999	Over half of a sample of 21 abstracts of accepted articles had deficiencies before the initiative and this dropped to zero out of 27 abstracts afterwards.
<b>Providing instructions to authors</b>	
Asano 1995b RA	After a requirement for authors to supply the first page of each reference cited, citation error dropped from 48% (45/94) in 1990 to 22% (21/96) in 1994
Fister 2005	Small but statistically significant improvements in completely accurate and technically correct references were seen in the instructional group, or the brief reminder group compared with standard practice. No significant differences were seen for substantive errors (standard practice 437/720 references (61%); brief reminder 311/613 (51%); instructional 365/702 (52%)).
Jackson 2003 RA	A significant improvement in citation accuracy from 1985 (30 incorrect references out of 100) to 1995 (11 incorrect references out of 100) is attributed to requiring authors to submit first pages of all references cited in their manuscripts

**study results** (Continued)

Karlawish 1999	Quality of reporting was assessed by using four measures identified from publications outlining research ethics requirements. Reporting of ethics requirements ranged from all 45 papers (100%) reporting their study justification, 36 (80%) papers reporting that informed consent had been obtained (or waived), 18 papers (40%) reporting Institutional Review Board review and 6 (13%) reporting nursing home committee review. For articles published in journals giving no instructions (n=9) the average quality score (out of 4) was 1.4; for the group with instructions less than Uniform requirements (n=7) the average quality score was 2.5, for the group with instructions conforming to the Uniform requirements (n=24) the quality score was 2.4 and for the group conforming to Uniform requirements plus giving additional instructions (n=5) the quality score was 3.2 (Kruskal-Wallis chi-square = 11.2, p = 0.01).
Nishina 1995c RA	After authors were instructed to consult original sources for references, citation error dropped from 44% (42/96) in 1990 to 29% (28/97) in 1994
Nishina 2000 RA	After authors were instructed to consult original sources for references, citation error dropped from 1990 to 26% (25/98) in 1998 and 27% (26/97) in 1999
Pitkin 1998	The types of defects in the 55 defective abstracts were: - inconsistencies between the body of the paper and the abstract (51% of total errors, 95% CI 38% to 64%; n=28) - data in the abstract but not in the body of the paper (29%, 95% CI 17% to 41%; n=15) - both the above defects (15%, 95% CI 10% to 20%; n=8) - unjustified conclusions in the abstract (5%, 95% CI 3 to 7; n=3). Pitkin also surveyed a small sample of 1995 and 1996 issues of four journals for defects in abstracts. The percentage of defective abstracts ranged from 27% to 65% but the investigators did not attempt to identify the cause of this wide range: New England Journal of Medicine; 27% (3 deficient out of 11 abstracts) JAMA; 50% (7 out of 14) American Journal of Obstetrics and Gynecology; 53% (19 out of 36) Pediatrics; 65% (13 out of 20).
<b>Providing instructions to readers</b>	
Gross 1994	With an example, 83% of observations (33/40) identified the correct model compared with 86% (36/42) for readers' observations without an example. For ability to derive correct values, the corresponding figures were 88% (35/40) and 57% (24/42).
<b>Structuring abstracts</b>	
Booth 1997	Overall searching precision (percentage of references retrieved which were relevant) for ten searches in a simulated database was 45% for structured abstracts and 42% for unstructured abstracts. Search precision was better with structured abstracts than unstructured in five of the ten searches, the same in one search and worse in four searches. Overall searching recall (percentage of 'gold standard' (i.e. all relevant) references retrieved) for ten searches was 32% for structured abstracts and 75% for unstructured abstracts. Recall of structured abstracts was worse in nine of the ten searches and the same for one search.

**study results** (Continued)

Comans 1990	<p>Although structured abstracts (n=15) from Annals of Internal Medicine, BMJ and New England Journal of Medicine were judged to be clear and detailed, they often had the following information missing:</p> <ul style="list-style-type: none"> <li>- sociodemographic features of patients</li> <li>- patient selection methods</li> <li>- methods of statistical analysis</li> </ul> <p>Unstructured abstracts (n=21) from Nederlands Tijdschrift voor Geneeskunde often had the following information missing:</p> <ul style="list-style-type: none"> <li>- details of objective</li> <li>- setting of the study</li> <li>- sociodemographic features of patients and other patient details</li> <li>- details of methods</li> </ul>
Dupuy 2003	<p>In a comparison of abstracts of clinical studies published in 2000 in 3 dermatology journals (Archives of Dermatology, British Journal of Dermatology and the Journal of the American Academy of Dermatology), structured abstracts (n=34) scored significantly better than unstructured abstracts (n=15): 0.71 (SD0.11) versus 0.56 (SD0.18); p=0.002.</p> <p>Structured abstracts were longer on average than unstructured abstracts: 256 words (SD77) versus 169 (SD65); p&lt;0.001.</p> <p>A strong positive correlation between length and score was observed for unstructured abstracts (Pearson correlation coefficient 0.75; p=0.002) while no such significant correlation was seen for structured abstracts (Pearson correlation coefficient 0.30, p=0.08)</p>
Harbourt 1995	<p>All 924,478 MEDLINE records for 1989-1991 were compared with the subset of 3873 records with structured abstracts</p> <p>MeSH:</p> <p>Average of 3 more headings in structured abstracts than in MEDLINE records overall (14.1 structured versus 10.1 overall)</p> <p>Clinical trials: mean of 15.3 headings for structured abstracts (n=581 records) versus overall mean of 13.2 (n=18,495 records)</p> <p>Reviews: mean of 10.1 headings for structured abstracts (n=116 records) versus overall mean of 8.2 (n=92,475 records)</p> <p>Abstract length (n's as for MeSH):</p> <p>Average length of a structured abstract is approximately 700 characters longer than the overall average (1,739.2 structured versus 1,062.8 overall)</p> <p>Clinical trials: mean of 1,826.9 characters for structured abstracts versus overall mean of 1,195.0</p> <p>Reviews: mean of 1,749.1 characters for structured abstracts versus overall mean of 977.3</p>
Hartley 1996a	<p>30 pairs of unstructured and structured (rewritten) abstracts from the British Journal of Educational Psychology were compared for time taken to search for information from the abstracts - readers searched significantly faster and made significantly fewer errors when using structured abstracts.</p>
Hartley 1996b	<p>In a companion study to Hartley 1996a, readers also searched significantly faster and made significantly fewer errors when using structured abstracts, although there was a 'learning' effect apparent in those readers who were allocated structured abstracts before unstructured ones.</p>

study results (Continued)

Hartley 1996c	Over 400 readers stated their preferences for different versions of an abstract which was modified in regard to typography, layout and position on the page. The most preferred version was bold capital letters for subheadings, a line-space above the main heading and centring of the abstract over the top of a the subsequent two-column article.
Hartley 1997	The readability scores of BMJ and British Journal of Psychiatry (BJP) abstracts published before (20 abstracts from each journal) and after (20 abstracts from each journal) the introduction of structured abstracts showed no significant difference in either the Flesch Reading Ease or Gunning Fog Index (BMJ Flesch t test (one tailed) = 0.12, p = ns, BMJ Gunning Fog 1.03, p = ns; BJP Flesch 0.40, p = ns, Gunning Fog 0.98, p = ns). However abstract length (number of words) was significantly greater in the structured abstracts (BMJ t test (one tailed) = 3.20, p<0.0005; BJP 2.64 p<0.01). When a single editor rewrote 30 unstructured abstracts as structured abstracts, the readability scores were significantly improved, and the abstract length significantly increased (Flesch t test 4.47, p<0.0005; Gunning Fog 2.62, p<0.01; abstract length 5.90, p<0.0005). These results were consistent when 29 unstructured abstracts were rewritten by the original 29 authors (Flesch t test 2.09, p<0.05; Gunning Fog 3.25, p<0.005, abstract length 2.20, p<0.025). When 108 readers were asked to put scrambled sentences of an abstract (with the headings removed) in order, they made fewer errors with structured abstracts (mean 0.69 SD 0.98) than with unstructured ones (mean number of errors 3.40 SD 2.01); t test (two-tailed)8.85, p<0.001. However another study involving student readers and some differences in how the information was scrambled, did not show differences in most structured and unstructured abstract comparisons. Sixty-three readers rated the structured version of a single abstract easier to read on a subjective 10 point scale, compared with the unstructured version (correlated t = 4.89, df 62, p<0.001, two tail test). The mean score was 6.10 (SD 2.01) for the unstructured version and 7.92 (SD 1.83) for the structured version of the abstract.
Hartley 1998	A checklist (based on Taddio 1994) and intended to measure the information content of the abstracts also showed improved scores for the structured versions of the abstracts, the mean score for unstructured abstracts was 6.4 (SD 2.8) out of a possible top score of 22, and the mean score for the structured version of the abstract was 9.1 (SD 2.6), t = 6.04, p (one-tailed<0.0005) . A crude measure suggests that student evaluators took about four minutes to evaluate each unstructured abstract and about three minutes for each structured abstract.
Hartley 2000	30 unstructured abstracts for papers submitted to journals published by the British Psychological Society rewritten as structured abstracts: very similar with regard to accuracy (few inaccuracies in either set of abstracts)
Hartley 2002	When the length of unstructured abstracts was increased or the length of structured abstracts decreased in 15 journals, pagination of articles was not usually affected, except where the journal's pagination policy is start a new article on the same page as a previous article (a format rarely used in scientific journals)
Hartley 2003	24 unstructured abstracts from the Journal of Educational Psychology rewritten as structured abstracts: Abstract length, mean: Structured 186 words [SD 15] versus unstructured 133 [SD 22], p< 0.001 Sentence lengths, mean: Structured 20.8 words [SD 3.0] versus unstructured 24.6 [SD 8.3], p < 0.02 Percentage of passives, mean: Structured 23.7 [SD 17.3] versus unstructured 32.7 [SD 22.8], pns

study results (Continued)

	<p>Flesch reading score, mean: Structured 31.1 [SD 12.1] versus unstructured 21.1 [SD 13.7], <math>p &lt; 0.001</math></p> <p>Use of longer words, mean score: Structured 35.8 [SD 4.6] versus unstructured 40.0 [SD 5.3], <math>p &lt; 0.001</math></p> <p>Use of common words, mean score: Structured 61.1 [SD 6.3] versus unstructured 57.7 [SD 8.6], <math>p &lt; 0.01</math></p> <p>Use of present tense, mean: Structured 4.1 [SD 1.9] versus unstructured 2.7 [2.8], <math>p &lt; 0.01</math></p> <p>Information checklist, mean score: Structured 9.7 [SD 1.4] versus unstructured 5.5 [SD 1.0], <math>p &lt; 0.001</math></p> <p>Clarity ratings, mean: Structured 7.4 [SD 2.0] versus 6.2 [2.0], <math>p &lt; 0.01</math></p>
Khosrotehrani 2002	<p>Assessed abstract quality in <i>Annales de Dermatologie</i> before and after the introduction of structured abstracts in 1993:</p> <p>Mean scores (based on Narine):</p> <p>1991-92: 0.72 (SD -0.20), <math>n=8</math></p> <p>1996: 0.69 (SD -0.12) <math>n=17</math></p> <p>2000: 0.83 (SD -0.08) <math>n=18</math></p> <p>Nonsignificant trend towards improved scores, reported as <math>p = 0.015</math>; should be 0.15?</p>
Scherer 1998	<p>A comparison of unstructured and structured abstracts in the Archives of Ophthalmology showed an improved CONSORT abstract 'score' (maximum score = 9) for the structured abstracts (structured mean 6.8 (standard error of the mean (SEM) 0.7), <math>n=9</math>; unstructured mean 4.6 (SEM 0.4) <math>n=17</math>, <math>p=0.008</math>). However no statistically significant difference in this score was seen for structured abstracts compared with unstructured abstracts in Ophthalmology (structured mean 5.6 (SEM 0.3) <math>n=28</math>; unstructured mean 4.9 (SEM 0.4), <math>n=23</math>). No statistically significant difference was seen for either journal when the CONSORT criteria were scored across the text of the paper rather than just the abstract, and no difference was seen over time (1991/92 compared to 1993/94) for unstructured abstracts in both journals. No statistically significant increase in CONSORT 'score' of the text was seen in either the Archives of Ophthalmology (structured mean score 12.3 (SEM 1.3) <math>n=9</math>; unstructured mean score 15.7 (SEM 1.1) <math>n=17</math>) or Ophthalmology (structured mean score 16.9 (SEM 0.8) <math>n=28</math>; unstructured mean score 16.0 (SEM 0.9) <math>n = 23</math>) when papers with structured abstracts were compared to papers with unstructured abstracts. The authors comment that "reporting of the CONSORT criteria in the text was unimpressive".</p>
Taddio 1994	<p>A comparison of 150 unstructured and 150 structured abstracts in three journals (BMJ, JAMA and CMAJ) showed the structured abstracts to be of higher quality, as measured by 33 objective criteria (unstructured mean score 0.57, structured mean score 0.74, <math>p&lt;0.001</math>). Quality scores did not show a statistically significantly difference between years (1988 and 1989) or between journals except for the comparison between unstructured abstracts in BMJ and JAMA, with a lower score for BMJ abstracts, <math>p&lt;0.05</math>. Two journals provided detailed instructions on how to write an abstract while one did not.</p>
Trakas 1997	<p>Statistically significant improvement in the quality of structured abstracts compared to unstructured abstracts, as measured by a checklist of 29 objective criteria (structured mean score 62.5 out of a possible 100 (SD 11.0); unstructured mean score 53.3 (SD 10.0), <math>F = 9.48</math>, <math>p = 0.03</math>). No statistically significant difference was detected between journal types (pharmacy, medical or health economics) or between years (1990, 1991, 1992, 1993, 1994). There was a correlation between the subjective scores given by experienced raters and the quality of abstracts as measured by the set of objective criteria</p>

**study results** (Continued)

Wilczynski 1995	Many search terms were comparable for structured and unstructured abstracts, but some performed better in MEDLINE with structured abstracts, particularly for aetiology and prognosis articles
Wong 2005	Structured abstracts (1991/2 and 2001/2) were of higher quality than unstructured abstracts from 1988/89 issues of the same journals; but no significant improvement in abstract quality was seen between 1991/2 and 2001/2

**Analysis 2.1. Comparison 2 Citation and quotation accuracy, Outcome 1 Error rates (proportion of incorrect references).**

**Error rates (proportion of incorrect references)**

Acea 1997 RA	Nebril	56/91 (62%)	3 major errors			71 total errors	gastro-intestinal medicine
Aronsky RA	2005	225/656 (34%)				311 total errors	medical informatics
Asano 1995a RA		1990: 31/98 (32%) 1994: 41/99 (41%)	6 major errors			80 total errors	anaesthesia
Asano 1995b RA		1990: 45/94 (48%) 1994: 21/96 (22%)	1990: 5% 1994: 3%			1990: 63 total errors 1994: 24 total errors	anaesthesia
Avila 1996 RA		54/100 (54%)	12 major errors				anaesthesia
Browne RA	2004	145/259 (56%)				submitted papers	radiology
Buchan RA	2005	32/200 (16%)		50/200 (25%)	30 not accurate and 20 partially accurate	35 total citation errors	ophthalmology
Cakir 2003 RA		117/182 (64%)	7 major errors				orthopaedics
Celayir 2003 RA		443/1312 (34%)				520 total errors	paediatric surgery
de Lacey RA	1985			45/300 (15%)		71 citation errors in 300 references (9% major): overall - range of 8% to 46% be-	general medicine

**Error rates (proportion of incorrect references)** (Continued)

					tween 6 journals	
Doms 1989 RA	211/500 (42%)	73 major errors			Citation error: range of 37% to 49% between 5 journals: 249 total errors	dentistry
Eichorn 1987 RA	46/150 (31%)	5 major errors	45/150 (30%)	23 major errors		public health
Evans 1990 RA	54/150 (36%)	13 total errors	40/150 (27%)	37 major errors		surgery
Fenton 2000 RA	63/168 (38%)	20 major errors	26/153 (17%)	18 major errors		otolaryngology/head and neck surgery
Ferreira 2000 RA	91/223 (41%)	60 major out of 162 total errors (37%)				obstetrics and gynaecology
Foreman 1987 RA	35/112 (31%)	3 major errors				nursing
George 1994 RA	99/240 (41%)		83/239 (35%)	34 major errors	Only 36% of references were completely accurate (both citation and quotation correct) Relationship between rate of citation errors and journals which monitor citations was of "borderline significance" p = 0.066	dermatology
Goldberg 1993 RA	40/145 (28%)	16 major errors	Qualitative: 51/145 (35%) Quantitative: 8/17 (47%)	30% (qualitative)		emergency medicine
Goodrich 1977 RA	634/2195 (29%)				ranged from 14% to 50% between 10 jour-	general medicine

**Error rates (proportion of incorrect references)** (Continued)

					nals		
Gosling RA	2004	115/320 (36%)		70/565 (12%)		121/160 (76%) major errors (which includes multiple errors in references)	manual therapy
Gupta RA	2005	69/176 (39%)		15/176 (9%)			paediatrics
Hansen RA	1994	34/95 (36%)	3 major errors	9/95 (10%)	7 major errors	35 citation total errors	radiology
Hobma RA	1992	31/100 (31%)	5 major errors	44/100 (44%)			general medicine
Holt RA	2000	425/1022 (42%)				754 total errors	general medicine
Jackson RA	2003	1985: 30/100 (30%) [33 total errors] 1995: 11/100 (11%) [12 total errors]	1985: 2 1995: 2			Required copy of first page of each reference from 1995	hand surgery
Key RA	1977	1005/1867 (54%)				Further 6% (115) of references could not be verified	physical medicine and rehabilitation
Kolbitsch RA	1997			4/32 (13%) contradictions 15/32 (47%) selective quotations			anaesthesia
Lawson RA	1999		5 major errors	10/147 (7%)	8 major errors	total of 56 citation errors in 147 references	psychiatry
Lee RA	1999	57/200 (29%)	3 major errors	41/200 (21%)	23 major errors		dermatology
Lok RA	2001	240/550 (44%)				incidence of citation errors negatively correlated with the journal impact factor (p=0.02) and im-	nursing

**Error rates (proportion of incorrect references)** (Continued)

					mediacy index (p=0.03)	
Lowry 1985 RA	20/248 (8%)		20/61 (33%)	7 major errors	also see Table 1.1	BMJ
Lukic 2004 RA	54/199 (38%)		52/272 (19%)			anatomy
McLellan 1992 RA	175/348 (50%)				No statistically significant differences seen between 4 journals	anaesthesia
Mikawa 1996 RA	40/94 (43%)	6 major errors			total of 56 errors	intensive care
Neihouse 1999 RA			31/100 (31%)			pharmacology
Ngan Kee 1997a RA	54/90 (60%)					surgery
Ngan Kee 1997b RA	112/200 (56%)				total of 152 errors	Hong Kong Medical Journal
Nishina 1995a RA	38/96 (40%)	5 major errors				intensive care
Nishina 1995b RA	1990: 34/95 (36%) 1994: 36/96 (38%)					anaesthesia
Nishina 1995c RA	1990: 42/96 (44%) 1994: 28/97 (29%)				Statistically significant difference (p<0.05) between 1990 and 1994	anaesthesia
Nishina 1995d RA	1990: 52/98 (53%) 1993: 44/97 (45%)				Not a statistically significant difference between 1990 and 1993	anaesthesia
Nishina 1995f RA	1987: 38/92 (41%) 1994: 39/93 (42%)	2 2			Not a statistically significant difference between 1987 and 1994	anaesthesia

**Error rates (proportion of incorrect references)** (Continued)

Nishina RA	2000	1998: 25/98 (26%) 1999: 26/97 (27%)				Statistically significant difference between 1990 and 1998, and 1999	anaesthesia
Nuckles RA	1993	64/298 (22%)					dentistry
O'Connor RA	2002	35/100 (35%)				total of 41 errors	emergency medicine
Oermann RA	2001	79/190 (42%)	29%				paediatric nursing
Oermann RA	2002a	56/244 (23%)	20%				critical care nursing
Oermann RA	2002b	33/130 (25%)	19%			Journal with in-house librarian to check references had few citation errors	general nursing
Oermann RA	2002c	54/221 (24%)	21%				maternal and neonatal nursing
Orlin RA	1996	123/472 (26%)				153 errors overall - 32 (21%) of these made the reference "inconvenient" to locate	oral and maxillo-facial surgery
Perez Garcia RA	2000	189/433 (44%)	14%				nephrology
Pieters RA	2001	42/100 (42%)	10 major errors	15/100 (15%)	3 major errors		psychiatry
Pulida RA	1995	236/368 (64%)	94 major errors			multiple errors in 45% of references; no trends over time (1962 to 1992) detected	general medicine
Putterman RA	1991	128/384 (33%)	20 major errors			total of 136 errors	general medicine

**Error rates (proportion of incorrect references)** (Continued)

Putterman 1992 RA			27/120 (23%)	8 major errors		general medicine
Roach 1997 RA	82/133 (62%)				No statistically significant differences found between 3 journals	obstetrics and gynaecology
Schulmeister 1998 RA	58/180 (32%)	43 major errors	12/180 (7%)		Statistically significant differences in error rates of the 3 journals, p=0.039	nursing
Siebers 1999 RA	43/99 (43%)	27 major errors			total of 64 errors	medical laboratory science
Siebers 2000a RA	521/1787 (29%)				total of 754 errors	allergy
Siebers 2000b RA	300/1557 (19%)					general medicine (5 leading medical journals)
Siebers 2001 RA	226/892 (25%)	32 major errors			total of 341 errors	clinical chemistry
Sutherland 2000 RA	see comments	14 major errors			only reports total numbers of errors - total of 122 errors in 400 references (3 to 37 per 100 references), not how many references contained errors	orthopaedics
Taylor 1998 RA	120/262 (46%)				total of 148 errors	nursing
VargasOrigel 2001 RA	119/400 (30%)	8 major errors			total of 119 errors in 400 references	paediatrics
Warren 1997 RA	63/240 (26%) major journals 49/142 (35%)		36/240 (15%) major journals 28/142 (20%)		combined error rate of references from the minor	infectious diseases

**Error rates (proportion of incorrect references)** (Continued)

	minor journals		minor journals		journals was significantly greater than major journals (p=0.059)	
--	----------------	--	----------------	--	--	--

**APPENDICES**

**Appendix I. Search strategy - CMR, MEDLINE AND EMBASE**

*Cochrane Methodology Register, last searched on Cochrane Library Issue 2, 2007*

- editing, as a text word

We also scanned all of the titles (and abstracts, where appropriate) in the Register.

**MEDLINE (from inception to July 2006); OVID platform**

- Writing as a text word, last searched July 2006
- Copyediting or copy-editing as text words last searched July 2006
- (accuracy or accurate or error\$ or inaccurate or inaccuracy) and (reference\$ or citation\$ or quotation\$) as text words, last searched July 2006
- Proofreading or proof reading or proof-reading, as text words, last searched July 2006
- Editing, as a text word, last searched July 2005

**EMBASE (from inception to June 2007); OVID platform**

- Writing as an EMTREE heading (exploded), last searched June 2007
- Copy-editing or copyediting as text words, last searched June 2007
- Redaction as a text word, last searched June 2006
- Editing (text word) and (journal or literature; text words), last searched June 2007
- Proofreading or proof reading or proof-reading as text words, last searched September 2006
- Medical literature and Quality control as EMTREE headings (focused), last searched September 2006

**WHAT'S NEW**

Last assessed as up-to-date: 25 July 2007.

13 August 2008	New citation required but conclusions have not changed	New search was conducted with the addition of new studies. Structural changes were also made to the results tables.
26 July 2007	New search has been performed	This review has been updated (new search June 2007) from a previously published review (Wager 2003). The following studies were added: Technical editing (14 new studies, making a total of 32

(Continued)

		<p>studies).</p> <p><i>New studies:</i> Dupuy 2003; Fister 2005; Harbourn 1995; Hartley 1996a; Hartley 1996b; Hartley 1996c; Hartley 2000; Hartley 2002; Hartley 2003; Khosrotehrani 2002; Siegel 2005; Silagy 1998; Wilczynski 1995; Wong 2005</p> <p>Reference accuracy (31 new studies, making a total of 66 studies).</p> <p><i>New studies:</i> Acea Nebril 1997; Aronsky 2005; Browne 2004; Buchan 2005; Cakir 2003; Celayir 2003; Ferreira 2000; Gosling 2004; Gupta 2005; Jackson 2003*; Lawson 1999; Lok 2001; Lukic 2004; Ngan Kee 1997b; Nishina 1995f; Nishina 2000*; Nuckles 1993; O'Connor 2002; Oermann 2001; Oermann 2002a; Oermann 2002b*; Oermann 2002c; Orlin 1996; Perez Garcia 2000; Pieters 2001; Pulida 1995; Riesenber 2001; Siebers 2001; Sutherland 2000; Vargas-Origel 2001; Warren 1997</p> <p>* = reference accuracy studies also included in technical editing section</p> <p>Some structural changes were made to the results tables.</p>
--	--	--

## HISTORY

Protocol first published: Issue 2, 2001

Review first published: Issue 1, 2003

---

27 December 2007	Amended	Converted to new review format.
------------------	---------	---------------------------------

---

## CONTRIBUTIONS OF AUTHORS

For the first version of this review, both review authors contributed equally to the preparation of the review.

For the 2007 update, Philippa Middleton took the lead on integrating the new technical editing studies, while Elizabeth Wager took the lead in preparing the reference accuracy component. Both review authors contributed equally to the drafting of the updated discussion, interpretation and conclusions.

## **DECLARATIONS OF INTEREST**

The reviewers publish papers in peer-reviewed biomedical journals and some of their work involves technical editing.

One of the review authors (PM) was a co-investigator on one of the studies ([Silagy 1998](#)) included in this review.

## **SOURCES OF SUPPORT**

### **Internal sources**

- NHS Research and Development Programme, UK.
- Glaxo Wellcome Research and Development, UK.

### **External sources**

- No sources of support supplied

## **INDEX TERMS**

### **Medical Subject Headings (MeSH)**

Editorial Policies; Peer Review, Research [standards]; Periodicals as Topic [\*standards]; Publishing [\*standards]